PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : G01W 33/68 — A2

(11) International Publication Number: **WO 00/07038**

(43) International Publication Date: 10 February 2000 (10.02.00)

(21) International Application Number: PCT/CA99/00702

(22) International Filing Date: 30 July 1999 (30.07.99)

(30) Priority Data:

2,244,349	30 July 1998 (30.07.98)	CA
09/124,850	30 July 1998 (30.07.98)	US

(71) Applicant *(for all designated States except US)*: UNIVERSITE DE MONTREAL [CA/CA]; Postal Code 6128, Station A, Montreal, Quebec H3C 3J7 (CA).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: MICHNICK, Stephen, William, Watson [CA/CA]; 330 Grosvenor, Westmount, Quebec H3C 2M2 (CA). PELLETIER, Joelle, Nina [CA/CA]; Apartment #611, 10 Rosemount Avenue, Westmount, Quebec H3Y 3K4 (CA). REMY, Ingrid [CA/CA]; 8526 Berri, Montreal, Quebec H2P 2G4 (CA).

(74) Agents: DUBUC, Jean, H. et al.; Goudreau Gage Dubuc & Martineau Walker, The Stock Exchange Tower, Suite 3400, 800 Place Victoria, P.O. Box 242, Montreal, Quebec H4Z 1E9 (CA).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: PROTEIN FRAGMENT COMPLEMENTATION ASSAYS FOR THE DETECTION OF BIOLOGICAL OR DRUG INTERACTIONS

(57) **Abstract**

A strategy for designing and implementing protein–fragment complementation assays (PCAs) to detect biomolecular interactions *in vivo* and *in vitro* is described herein. A Protein Complementation Assay/Universal Reporter System (PCA/URS) for detecting and screening for agonists and antagonists of a membrane receptor is also described. The design, implementation and broad applications of this strategy are illustrated with a large number of enzymes with particular detail provided for the example of murine dihydrofolate reductase (DHFR). Fusion peptides consisting of *N*–and *C*–terminal fragments of murine DHFR fused to GCN4 leucine zipper sequences were coexpressed in *Escherichia coli* grown in minimal medium, where the endogenous DHFR activity was inhibited with trimethoprim. Coexpression of the complementary fusion products restored colony formation. Survival only occurred when both DHFR fragments were present and contained leucine–zipper forming sequences, demonstrating that reconstitution of enzyme activity requires assistance of leucine zipper formation. DHFR fragment–interface point mutants of increasing severity (Ile to Val, Ala and Gly) resulted in a sequential increase in *E. coli* doubling times illustrating the successful DHFR fragment reassembly rather than non–specific interactions between fragments. This assay could be used to study equilibrium and kinetic aspects of molecular interactions including various types of interactions. The selection and design criteria applied here is developed for numerous examples of clonal selection, colorometric, fluorometric and other assays based on enzymes whose products can be measured. The development of such assay systems is shown to be simple, and provides for a diverse set of protein fragment complementation applications.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

TITLE OF THE INVENTION

PROTEIN FRAGMENT COMPLEMENTATION ASSAYS FOR THE DETECTION OF BIOLOGICAL OR DRUG INTERACTIONS

FIELD OF THE INVENTION

The present invention relates to the determination of the function of novel gene products. The invention further relates to Protein fragment Complementation Assays (PCA). PCAs allow for the detection of a wide variety of types of protein-protein, protein-RNA, protein-DNA, Protein-carbohydrate or protein-small organic molecule interactions in different cellular contexts appropriate to the study of such interactions.

BACKGROUND OF THE INVENTION

Many processes in biology, including transcription, translation, and metabolic or signal transduction pathways, are mediated by non-covalently-associated multienzyme complexes[1, 101]. The formation of multiprotein or protein-nucleic acid complexes produce the most efficient chemical machinery. Much of modern biological research is concerned with identifying proteins involved in cellular processes, determining their functions and how, when, and where they interact with other proteins involved in specific pathways. Further, with rapid advances in genome sequencing projects there is a need to develop strategies to define "protein linkage maps", detailed inventories of protein interactions that make up functional assemblies of proteins[2,3]. Despite the importance of understanding protein assembly in biological processes, there are few convenient methods for studying protein-protein interactions *in vivo*[4,5]. Approaches include the use of chemical crosslinking reagents and resonance energy transfer between dye-coupled proteins[102, 103]. A powerful and commonly used strategy, the yeast two-hybrid system, is used to identify novel protein-protein interactions and to examine the amino acid determinants of specific protein interactions[4,6-8]. The approach allows for rapid screening of a large number of clones, including cDNA libraries. Limitations of this technique includ the fact that the interaction must occur in a specific context (the nucleus of *S.*

cerevisiae), and generally cannot be used to distinguish induced versus constitutive interactions.

Recently, a novel strategy for detecting protein-protein interactions has been demonstrated by Johnsson and Varshavsky[108] called the ubiquitin-based split protein sensor (USPS)[9]. The strategy is based on cleavage of proteins with *N*-terminal fusions to ubiquitin by cytosolic proteases (ubiquitinases) that recognize its tertiary structure. The strategy depends on the reassembly of the tertiary structure of the protein ubiquitin from complementary *N*- and *C*-terminal fragments and crucially, on the augmentation of this reassembly by oligomerization domains fused to these fragments. Reassembly is detected as specific proteolysis of the assembled product by cytosolic proteases (ubiquitinases). The authors demonstrated that a fusion of a reporter protein-ubiquitin *C*-terminal fragment could also be cleaved by ubiquitinases, but only if co-expressed with an *N*-terminal fragment of ubiquitin that was complementary to the *C*-terminal fragment. The reconstitution of observable ubiquitinase activity only occurred if the *N*- and *C*-terminal fragments were bound through GCN4 leucine zippers[109,110]. The authors suggested that this "split-gene" strategy could be used as an *in vivo* assay of protein-protein interactions and analysis of protein assembly kinetics in cells. Unfortunately, this strategy requires additional cellular factors (in this case ubiquitinases) and the detection method does not lend itself to high-throughput screening of cDNA libraries.

Rossi, F., C. A. Charlton, and H. M. Blau (1997) Proc. Nat. Acad. Sci. (USA) 94, 8405-8410) have reported an assay based on the classical complementation of α and ω fragments of β-galactosidase (β-gal) and induction of complementation by induced oligomerization of the proteins FKBP12 and the mamalian target of rapamycin by rapamycin in transfected C2C12 myoblast cell lines. Reconstitution of b-gal activity is detected using substrate fluorescein di-β-D-galactopyranoside using several fluorecence detection assays. While this assay bears some resemblance to the present invention, there are several significant distinguishing differences. First, this particular complementation approach has been used for over thirty years in a vast number of applications

including the detection of protein-protein interactions. Krevolin, M. and D. Kates (1993) U.S. Patent No. 5,362,625) teaches the use of this complementation to detect protein-protein interactions. Also achievement of β-gal complementation in mamalian cells has previously been reported (Moosmann, P. and S. Rusconi (1996) Nucl. Acids Res. 24, 1171-1172).

As in the USPS, the yeast-two hybrid strategy requires additional cellular machinery for detection that exist only in specific cellular compartments. There is therefore a need for a detection system which uses the reconstitution of a specific enzyme activity from fragments as the assay itself, without the requirement for other proteins for the detection of the activity. Preferably, the assay would involve an oligomerization-assisted complementation of fragments of monomeric or multimeric enzymes that require no other proteins for the detection of their activity. Furthermore, if the structure of an enzyme were known it would be possible to design fragments of the enzyme to ensure that the reassembled fragments would be active and to introduce mutations to alter the stringency of detection of reassembly. However, knowledge of structure should not be a prerequesite to the design of complementing fragments. The flexibility allowed in the design of such an approach would make it applicable to situations where other detection systems may not be suitable.

Recent advances in human genomics research has led to rapid progress in the identification of novel genes. In applications to biological and pharmaceutical research, there is now the pressing need to determine the functions of novel gene products; for example, for genes shown to be involved in disease phenotypes. It is in addressing questions of function where genomics-based pharmaceutical research becomes bogged down. There is therefore the need for advances in the development of simple and automatable functional assays. A first step in defining the function of a novel gene is to determine its interactions with other gene products in an appropriate context; that is, since proteins make specific interactions with other proteins or other biopolymers as part of functional assemblies, an appropriate way to examine the

function of a novel gene is to determine its physical relationships with the products of other genes.

Screening techniques for protein interactions, such as the yeast "two-hybrid" system, have transformed molecular biology, but can only be used to study specific types of constitutively interacting proteins or interactions of proteins with other molecules, in narrowly defined cellular and compartmental contexts and require a complex cellular machinery (transcription) to work. To rationally screen for protein interactions within the context of a specific problem requires more flexible approaches. Specifically, assays that meet criteria necessary not only to detecting molecular interactions, but also to validating these interactions as specific and biologically relevant are required.

A list of assay characteristics that meet such criteria are as follows:

1) Allow for the detection of protein-protein, protein-DNA/RNA or protein-drug interactions *in vivo* or *in vitro*.

2) Allow for the detection of these interactions in appropriate contexts, such as within a specific organism, cell type, cellular compartment, or organelle.

3) Allow for the detection of induced *versus* constitutive protein-protein interactions (such as by a cell growth or inhibitory factor).

4) Allow for a distincton between specific and non-specific protein-protein interactions by controlling the sensitivity of the assay.

5) Allow for the detection of the kinetics of protein assembly in cells.

6) Allow for screening of cDNA, small organic molecule, or DNA or RNA libraries for molecular interactions.

The present description refers to a number of documents, the content of which is herein incorporated by reference.

SUMMARY OF THE INVENTION

The present invention seeks to provide the above-mentioned needs for which the prior art is silent. The present invention provides a general strategy for detecting protein interactions with other biopolymers including other proteins, nucleic acids, carbohydrates or for screening small molecule libraries

for compounds of potential therapeutic value. In a preferred embodiment, the instant invention seeks to provide an oligomerization-assisted complementation of fragments of monomeric enzymes that require no other proteins for the detection of their activity. In one such embodiment, a protein-fragment complementation assay (PCA) based on reconstitution of dihydrofolate reductase activity by complementation of defined fragments of the enzyme in *E. coli* is hereby provided. This assay requires no additional endogenous factors for detecting specific protein-protein interactions (i.e. leucine zipper interactions) and can be conveniently extended to screening cDNA, nucleic acid, small molecule or protein design libraries for molecular interactions. In addition, the assay can also be adapted to the detection of protein interactions in any cellular context or compartment and be used to distinguish between induced *versus* constitutive protein interactions in both prokaryotic and eukaryotic systems.

The individual PCAs presented here are completely *de novo* designed interaction detection assays, not described in any way previously except for publications arising from applicants laboratory. Secondly, this application describes a general strategy to develop molecular interaction assays from a large number of enzyme or protein detectors, all *de novo* designed assays, whereas the β-gal assay is not novel. Thirdly, there are no general strategies or advancements over previously well documented applications given in the art.

One particular strategy for designing a protein complementation assay (PCA) is based on using the following characteristics: 1) A protein or enzyme that is relatively small and monomeric, 2) for which there is a large literature of structural and functional information, 3) for which simple assays exist for the reconstitution of the protein or activity of the enzyme, both *in vivo* and *in vitro*, and 4) for which overexpression in eukaryotic and prokaryotic cells has been demonstrated. If these criteria are met, the structure of the enzyme is used to decide the best position in the polypeptide chain to split the gene in two, based on the following criteria: 1) The fragments should result in subdomains of continuous polypeptide; that is, the resulting fragments will not disrupt the subdomain structure of the protein, 2) the catalytic and cofactor

binding sites should all be contained in one fragment, and 3) resulting new *N*- and *C*-termini should be on the same face of the protein to avoid the need for long peptide linkers and allow for studies of orientation-dependence of protein binding.

It should be understood that the above mentioned criteria do not all need to be satisfied for a proper working of the present invention. It is an advantage that the enzyme be small, preferably between 10-40 kDa. Although monomeric enzymes are preferred, multimeric enzymes can also be envisaged as within the scope of the present invention. The dimeric protein tyrosinase can be used in the instant assay. The information on the structure of the enzyme provides an additional advantage in designing the PCA, but is not necessary. Indeed, an additional strategy, to develop PCAs is presented, based on a combination of exonuclease digestion-generated protein fragements followed by directed protein evolution in application to the enzyme aminoglycoside kinase. Although the overexpression in prokaryotic cells is preferred it is not a necessity. It will be understood to the skilled artisan that the enzyme catalytic site (of the chosen enzyme) does not absolutely need to be on same molecule.

The present application explains the rationale and criteria for using a particular enzyme in a PCA. For PCA, a gene for a protein or enzyme is rationally dissected into two or more fragments. Using molecular biology techniques, the chosen fragments are subcloned, and to the 5' ends of each, proteins that either are known or thought to interact are fused. Co-transfection or transformation these DNA constructs into cells is then carried out. Reassembly of the probe protein or enzyme from its fragments is catalyzed by the binding of the test proteins to each other, and reconstitution is observed with some assay. It is crucial to understand that these assays will only work if the fused, interacting proteins catalyze the reassembly of the enzyme. That is, observation of reconstituted enzyme activity must be a measure of the interaction of the fused proteins.

A preferred embodiment of the present invention focuses on a PCA based on the enzyme dihydrofolate reductase. Expansion of the

strategy to include assays in eukaryotic, cells, library screening, and a specific application to problems concerning the study of integrated biochemical pathways such as signal transduction pathways, is presented. Additional assays, including those based on enzymes that can act as dominant or recessive drug selection or metabolic salvage pathways are disclosed. In addition, PCAs based on enzymes that will produce a colored or fluorescent product are also disclosed. The present invention teaches how the PCA strategy can be both generalized and automated for functional testing of novel genes, screening of natural products or compound libraries for pharmacological activity and identification of novel gene products that interact with DNA, RNA or carbohydrates. It also teaches how the PCA strategy can be applied to identifying natural products or small molecules from compound libraries of potential therapeutic value that can inhibit or activate such molecular interactions and how enzyme substrates and small molecule inhibitors of enzymes can be identified. Finally, it teaches how the PCA strategy can be used to perform protein engineering experiments that could lead to designed enzymes with industrial applications or peptides with biological activity.

Simple strategies to design and implement assays for detecting protein interactions *in vivo* are disclosed herein. Complementary fragments of the native mDHFR have been designed such that, when coexpressed in *E. coli* grown in minimal medium, they allow for survival of clones expressing the two fragments, where the basal activity of the endogenous bacterial DHFR is inhibited by the competitive inhibitor trimethoprim. Reconstitution of activity only occurred when both *N*- and *C*-terminal fragments of DHFR were coexpressed as C-terminal fusions to GCN4 leucine zipper sequences, indicating that reassembly of the fragments requires formation of a leucine zipper between the *N*- and *C*-terminal fusion peptides. The sequential increase in cell doubling times resulting from the destabilizing mutations directed at the assembly interface (Ile114 to Val, Ala or Gly) demonstrates that the observed cell survival under selective conditions is a result of the specific, leucine-zipper-assisted association of mDHFR

fragment[1,2] with fragment[3], as opposed to nonspecific interactions of Z-F[3] with Z-F[1,2]. Several detailed and many additional examples are given.

As demonstrated previously with the ubiquitin-based split protein sensor (USPS)[9], a protein-fragment complementation strategy can be used to study equilibrium and kinetic aspects of protein-protein interactions *in vivo*. The DHFR and other PCAs however, are simpler assays. They are complete systems; no additional endogenous factors are necessary and the results of complementation are observed directly, with no further manipulation. The *E. coli* cell survival assay described herein should therefore be particularly useful for screening cDNA libraries for protein-protein interactions. mDHFR expression in cells can be monitored by binding of fluorescent high-affinity substrate analogues for DHFR[26].

There are several further aspects of the PCAs that distinguish them from all other known strategies for studying protein-protein interactions *in vivo* (except USPS). Complementary fragments of enzymes that allow for controlling the stringency of the assay have been designed, and could be used to obtain estimates of the kinetics and equilibrium constants for association of two proteins. For example, with DHFR the point mutations of the wild-type enzyme Ile 114 to Val, Ala, or Gly alter the stringency of reconstitution of DHFR activity. For determining estimates of equilibrium and kinetic parameters for a specific protein-protein interaction, one could perform a series of DHFR PCA experiments with two proteins that interact with a known affinity, using the wild type or destabilizing mutant DHFR fragments. Comparison of cell growth rates in this model system with rates for a DHFR PCA using unknowns would give an estimate of the strength of the unknown interaction.

It should be understood that the present invention should not be limited to the DHFR or other PCAs presented herein, as they serve only as non-limiting embodiments of the protein complementation assay of the present invention. Moreover, the PCAs should not be limited in the context in which they could be used. Constructs could be designed for targeting the PCA fusions to specific compartments in the cell by addition of signaling peptide sequences[27,28]. Induced versus constitutive protein-protein interactions could be distinguished

by a eukaryotic version of the PCA, in the case of an interaction that is triggered by a biochemical event. Also, the system could be adapted for use in screening for novel, induced protein-molecular associations between a target protein and an expression library.

The instant invention is also directed to a method for detecting biomolecular interactions, the method comprising:

(a) selecting an appropriate reporter molecule;

(b) effecting fragmentation of the reporter molecule such that the fragmentation results in reversible loss of reporter function;

(c) fusing or attaching fragments of the reporter molecule separately to other molecules; followed by

(d) reassociation of the reporter fragments through interactions of the molecules that are fused to the fragments.

The invention also provides molecular fragment complementation assays for the detection of molecular interactions comprising a reassembly of separate fragments of a molecule, wherein reassembly of the fragments is operated by the interaction of molecular domains fused to each fragment of the molecules, and wherein reassembly of the fragments is independent of other molecular processes.

In another aspect, the present invention is directed to a method of testing biomolecular interactions comprising:

a) generating a first fusion product comprising

i) a first fragment of a first molecule; and

ii) a second molecule which is different or the same as the first molecule;

b) generating a second fusion product comprising

i) a second fragment of the first molecule; and

ii) a third molecule which is different from or the same as the first molecule or second molecule;

c) allowing the first and second fusion products to contact each other; and

d) testing for activity regained by association of the recombined fragments of the first molecule, wherein the reassociation is mediated by interaction of the second and third molecules.

In another novel feature, the invention is directed to a method comprising an assay where fragments of a first molecule are fused to a second molecule and fragment association is detected by reconstitution of the first molecule's activity.

The present invention also provides a composition comprising a product selected from the group consisting of:

(a) a first fusion product comprising:

i) a first fragment of a first molecule whose fragments can exhibit a detectable activity when associated and

ii) a second molecule that can bind (a)(i);

(b) a second fusion product comprising

i) a second fragment of the first molecule and

ii) a third molecule that can bind (b)(i); and

c) both (a) and (b).

The invention further provides a composition comprising complementary fragments of a first molecule, each fused to a separate fragment of a second molecule.

The inventors of the present subject matter further provide a composition comprising a nucleic acid molecule coding for a fusion product, which molecule comprises sequences coding for a product selected from the group consisting of:

(a) a first fusion product comprising:

i) fragments of a first molecule whose fragments can exhibit a detectable activity when associated and

ii) a second molecule fused to the fragment of the first molecule;

(b) a second fusion product comprising

i) a second fragment of the first molecule and

ii) a second or third molecule; and

c) both (a) and (b).

The present invention is also directed to a method of testing for biomolecular interactions associated with: (a) complementary fragments of a first molecule whose fragments can exhibit a detectable activity when

associated or (b) binding of two protein-protein interacting domains from a second or third molecule, the method comprising:

1) creating a fusion of

(a) a first fragment of a first molecule whose fragments can exhibit a detectable activity when associated and

(b) a first protein-protein interacting domain;

2) creating a fusion of

(a) a second fragment of the first molecule and

(b) a second protein-protein interacting domain that can bind the first protein-protein interacting domain;

3) allowing the fusions of (1) and (2) to contact each other; and

4) testing for the activity.

The instant invention further provides a composition comprising a product selected from the group consisting of:

(a) a first fusion product comprising:

i) a first fragment of a molecule whose fragments can exhibit a detectable activity when associated and

ii) a first protein-protein interacting domain;

(b) a second fusion product comprising

i) a second fragment of the first molecule and

ii) a second protein-protein interacting domain that can bind the first protein-protein interacting domain; and

(c) both (a) and (b).

The invention is also directed to a composition comprising a nucleic acid molecule coding for a fusion product, which molecule comprises sequences coding for either:

(a) a first fusion product comprising:

i) a first fragment of a molecule whose fragments can exhibit a detectable activity when associated and

ii) a first protein-protein interacting domain; or

(b) a second fusion product comprising

i) a second fragment of the molecule and

ii) a second protein-protein interacting domain that can bind the first protein-protein interacting domain; or

(c) both (a) and (b).

The invention also provides a method of detecting kinetics of protein assembly and screening cDNA libraries comprising performing PCA.

In another embodiment, the invention further provides a method of testing the ability of a compound to inhibit molecular interactions in a PCA comprising performing a PCA in the presence of the compound and correlating any inhibition with the presence.

In a further embodiment, the invention provides a method for detecting protein-protein interactions in living organisms and or cells, which method comprises:

(a) synthesizing probe protein fragments from an enzyme which enables dominant selection by dissecting the gene coding for the enzyme into at least two fragments;

(b) constructing fusion proteins with one or more molecules that are to be tested for interactions;

(c) fusing the proteins obtained in (b) with one or more of the probe fragments;

(d) coexpressing the fusion proteins; and

(e) detecting the reconstitution of enzyme activity.

The invention still provides a method for detecting biomolecular interactions, the method comprising:

(a) selecting an appropriate reporter molecule;

(b) effecting fragmentation of the reporter molecule;

(c) fusing or attaching fragments of the reporter molecule separately to other molecules; followed by

(d) reassociation of the reporter fragments through interactions of the molecules that are fused to the fragments.

The invention further relates to a method employing a Protein Complementation assay/Universal Reporter System (PCA/URS) for detecting and screening for agonists and antagonists of a membrane receptor, which method comprises:

a) generating a first nucleic acid vector encoding a first fusion product comprising:

i) a first fragment of a first PCA/URS reporter molecule, and

ii) a second molecule, fused to the first fragment, which comprises a first subdomain of a cellular receptor molecule of interest;

b) generating a second nucleic acid vector encoding a second fusion product comprising:

i) a second fragment of the first PCA/URS reporter molecule, and

ii) a third molecule, fused to the second fragment, which comprises a second subdomain of the cellular receptor, and where the second subdomain may be the same as the first subdomain in the case of a homodimeric cellular receptor, or different from the first subdomain in the case of a heterodimeric cellular receptor;

c) transfecting prokaryotic or eukaryotic cells with the first and second nucleic acid vectors;

d) testing the transfected cells for the PCA/URS reporter activity, the activity indicating reassociation of the first and second fragments of the PCA/URS molecule mediated by the interaction of the first and second subdomains of the cellular receptor molecule.

In a further embodiment, the invention is directed to a method employing a Protein Complementation Assay/Universal Reporter System (PCA/URS) for detecting and screening for agonists and antagonists of a membrane receptor, which method comprises:

a) generating a first nucleic acid vector encoding a first fusion product comprising:

i) a first fragment of a first PCA/URS reporter molecule, and

ii) a second molecule, fused to the first fragment, which comprises a first subdomain of a cellular receptor molecule of interest;

b) generating a second nucleic acid vector encoding a second fusion product comprising:

i) a second fragment of the first PCA/URS reporter molecule, and

ii) a third molecule, fused to the second fragment, which compris s a second subdomain of the cellular receptor, and where the second subdomain may be the same as the first subdomain in the case of a homodimeric cellular receptor, or different from the first subdomain in the case of a heterodimeric cellular receptor;

c) transfecting prokaryotic or eukaryotic cells with the first and second nucleic acid vectors;

d) obtaining a clonal population of cells that express the first and second fusion products; and

e) testing the transfected cells for the PCA/URS reporter activity, the activity indicating reassociation of the first and second fragments of the PCA/URS molecule mediated by the interaction of the first and second subdomains of the cellular receptor molecule.

In another embodiment, the invention relates to a method employing a Protein Complementation Assay/Universal Reporter System (PCA/URS) for detecting and screening for agonists and antagonists of a membrane receptor, which method comprises:

a) generating a first nucleic acid vector encoding a first fusion product comprising:

i) a first fragment of a first PCA/URS reporter molecule,

ii) a first linker, fused at one end to the first fragment, the linker region comprising between 1 and 30 amino acid residues; and iii) a second molecule, fused to the other end of the first linker, which comprises a first subdomain of a cellular receptor molecule of interest;

b) generating a second nucleic acid vector encoding a second fusion product comprising:

i) a second fragment of the first PCA/URS reporter molecule,

ii) a second linker, fused at one end to the second fragment, the linker comprising between 1 and 30 amino acid residue; and

iii) a third molecule, fused to the other end of the second linker, which comprises a second subdomain of the cellular receptor, and where the second subdomain may be the same as the first subdomain in the case of

a homodimeric cellular receptor, or different from the first subdomain in the case of a heterodimeric cellular receptor;

c) transfecting prokaryotic or eukaryotic cells with the first and second nucleic acid vectors;

d) testing the transfected cells for the PCA/URS reporter activity, the activity indicating reassociation of the first and second fragments of the PCA/URS molecule mediated by the interaction of the first and second subdomains of the cellular receptor molecule.

Lastly, the invention also provides a novel method of affecting gene therapy, which includes the step of providing the assays and compositions described above.

The present invention is pionneering as it is the first protein complementation assay displaying such a level of simplicity and versatility. The exemplified embodiments are protein-fragment complementation assays (PCA) based on mDHFR, where a leucine zipper directs the reconstitution of DHFR activity. Activity was detected by an *E. coli* survival assay which is both practical and inexpensive. This system illustrates the use of mDHFR fragment complementation in the detection of leucine zipper dimerization and could be applied to the detection of unknown, specific protein-molecular interactions *in vivo*.

It should be undertstood that the instant invention is not limited to the PCAs presented here, as numerous other enzymes can be selected and used in accordance with the teachings of the present invention. Examples of such markers can be found in Kaufman, (1987 Genetic Eng. 9:155-198) and references found therein as well as table 1 of this application.

It should also be clear to the skilled artisan to which the present invention pertains that the invention is not limited to the use of leucine zippers as the two interacting molecules. Indeed, numerous other types of protein-molecule interactions can be used and identified in accordance with the teaching of the present invention. The known types of motifs involved in protein-molecular interactions are well known in the art.

The present application refers to numerous prior art documents and the entire contents of all those prior art documents are herein incorporated by reference in their entirety.

Other features and advantages of the present invention will be apparent from the following description of the preferred embodiments thereof, the appended Examples and from the enjoined claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Having thus generally described the invention, reference will now be made to the accompanying drawings, showing by way of illustration a preferred embodiment thereof, and in which:

FIG. 1 provides a general description of a PCA. Using molecular biology techniques, the chosen fragments of the enzyme are subcloned, and to the 5' ends of each, proteins that either are known or thought to interact are fused. Co-transfection or transformation these DNA constructs into cells is then carried out and reconstitution with some assay is observed.

FIG. 2 is a scheme of the fusion constructs used in one of the embodiments of the invention. The hexahistidine peptide (6His), the homo-dimerizing GCN4 leucine zipper (Zipper) and mDHFR fragments (1, 2 and 3) are illustrated. The labels for the constructs are used to identify both the DNA constructs and the proteins expressed from these constructs.

FIG. 3: (A) shows *E. coli* survival assay on minimal medium plates. Control: Left side of the plate: *E. coli* harboring pQE-30 (no insert); right side: *E. coli* harboring pQE-16, coding for native mDHFR. Panel I: Left side of each plate: transformation with construct Z-F[1,2]; right side of each plate: transformation with construct Z-F[3]. Panel II: Cotransformation with constructs Z-F[1,2] and Z-F[3]. Panel III: Cotransformation with constructs Control-F[1,2] and Z-F[3]. All plates contain 0.5 mg/ml trimethoprim. In panels I to III, plates on the right side contain 1mM IPTG.

(B) *E. coli* survival assay using destabilizing DHFR mutants. Pan I I: Cotransformation of *E. coli* with constructs Z-F[1,2] and Z-F[3:Ile114Val]. Panel II: Cotransformation with Z-F[1,2] and Z-F[3:Ile114Ala].

Inset is a 5-fold enlargement of the right-side plate. Panel III: Cotransformation with Z-F[1,2] and Z-F[3:Ile114Gly]. All plates contain 0.5 mg/ml trimethoprim. Plates on the right side contain 1mM IPTG.

FIG. 4 features the coexpression of mDHFR fragments. (A) Agarose gel analysis of restriction pattern resulting from HincII digestion of plasmid DNA. Lane 1 contains DNA isolated from *E. coli* cotransformed with constructs Z-F[1,2] and Z-F[3]. Lanes 2 and 3 contain DNA isolated from *E. coli* transformed with, respectively, construct Z-F[3] and construct Z-F[1,2]. Fragment migration (in bp) is indicated to the right.

(B) SDS-PAGE analysis of mDHFR fragment expression. Lanes 1 to 5 show crude lysate of untransformed *E. coli* (lane 1), or *E. coli* expressing Z-F[1,2] (20.8 kDa; lane 2), Z-F[3] (18.4 kDa; lane 3), Control-F[1,2] (14.2 kDa; lane 4), and Z-F[1,2] + Z-F[3] (lane 5). Lane 6 shows 40 ml out of 2ml copurified Z-F[1,2] and Z-F[3]. Arrowheads point to the proteins of interest. Migration of molecular weight markers (in kDa) is indicated to the right.

FIG. 5 illustrates the general features of a PCA based on a survival assay such as the DHFR PCA. The assay can be used in a bacterial or a mammalian context. The inserted target DNA can be a known sequence coding for a protein (or protein domain) of interest, or can be a cDNA library.

FIG. 6 represents an autoradiograph of a COS cell lysate after a 30 min. ^{35}S-Met-Cys pulse-labelling. The expression pattern is essentially identical to that observed in *E. coli* (see Fig. 4). The DNA transfected into the cells (or cotransfected) is indicated above the respective lanes.

FIG. 7 illustrates the results of a protein engineering application of the mDHFR bacterial PCA. Two semi-random leucine zipper libraries were created (as described in the text) and each inserted *N*-terminal to one of the mDHFR fragments. Cotransformation of the resulting zipper-DHFR fragment libraries in *E. coli* and plating on selective medium allowed for survival of clones harboring successfully interacting leucine zippers. Fourteen clones were isolated and the zippers were sequenced to identify the residues at the "e" and "g" positions. The "e-g" pairs were categorized, as having attractive pairing (charge:charge, charge:neutral polar or neutral polar:neutral polar) or repulsive

pairing (charge:charge) and the number of each type of interaction scored for each clone. The total number of interactions for each clone is 6; the interactions are tallied on the histogram.

Fig. 8 is a schematic representation of Structure-based Epo receptor activation hypothesis (upper panel) and of the experimental strategy to test it (lower panel). (upper panel) Receptors are constitutive dimers in their unligated state. The extracellular domain exists in a conformation the holds the intracellular domains and associated JAK2 separated from each other by approximately 80 Å. On binding ligand (Epo or peptide agonist EMP1) the extracellular dimer is reorganized, bringing the intracellular domains to within 30 Å of each other, allowing autophosphorylation and activation of the JAK2s. (ii); The extracellular and transmembrane domains of murine EpoR are fused to one of two complementary fragments of murine DHFR (F[1,2] or F[3]) *via* flexible linkers (gray lines) consisting of $(Gly.Gly.Gly.Gly.Ser)_N$ repeats where N=1,2 or 6 to generate the following: EpoR-5aa-F[1,2] or -F[3], EpoR-10aa-F[1,2] or -F[3], EpoR-30aa-F[1,2] or -F[3] (See the legend of figure 9). Cells transfected with these fusions express receptors at the membrane surface. Fluorescein-methotrexate (fMTX) is taken up by cells and binds to reconstituted DHFR (F[1,2]+F[3]) and is retained in the cell. Unbound fMTX is rapidly released from the cells by active transport. Fusions in which DHFR fragments are connected to receptors by 5 or 10 amino acid linkers cannot or weakly complement in the inactive receptor (minimum separations or 40 or 80 Å, respectively). When receptor bind to Epo or EMP1 DHFR complementation can take place (separation 34 Å). (iii) Fusions with the 30aa linker allow complementation of DHFR fragments whether receptors are ligand bound or not. (iv) Results of complementation experiments with EpoR extracellular and transmembrane domain should be reproducible with complete EpoR receptor complex, including associated JAK2. Here is shown one such experiment in which DHFR fragments are fused to the C-terminal of JAK2 and co-expressed in cells along with full length EpoR.

Fig. 9 are the results of fluorescence microscopy of CHO DUKX-B11 cells expressing EpoR extracellular and transmembrane domains fused to DHFR complementary fragments (Constructs described in Fig. 8) and exposed to fMTX in the presence or absence of Epo or EMP1. All fusion clones were generated by PCR amplification of individual genes of interest. The construction of the DHFR F[1,2] and F[3] have been previously described. Oligonucleotides coding for Flexible linker peptides were synthesized individually with 5' and 3' complementary overhangs corresponding to 5' or 3' insertion between EpoR and DHFR fragment encoding sequences. regions of each construct were subcloned into the mammalian expression vector pMT3. Cells were stably lipofectamine (Life Technologies/ Gibco BRL) transfected with EpoR-DHFR fragment and stable colonies selected on alpha-MEM enriched with dialyzed 10 % fetal bovine serum (dialyzed to remove nucleotides, rendering cells dependent on exogenous DHFR activity) and in the presence of 2 nM human recombinant Epo (R. W. Johnson Pharmaceutical Research Institute). For microscopy, cells were grown on 18 mm glass cover slips to approximately 1×10^5 in 12 well plates. fMTX (Molecular Probes) was added to each sample at a final concentration of 10 iM and incubated for 22 hours at 37°C. Prior to microscopy, cells were treated with 10 nM Epo or 10 iM EMP1 for 30 minutes at 37°C. The medium was removed and the cells were washed with PBS (phosphate-buffered saline) extensively and reincubated for 15 minutes in alpha-MEM and Epo or EMP1 to allow for efflux of unbound fMTX. Medium was removed and cells were washed 4 times with PBS on ice and finally mounted on glass slides. Fluorescent microscopy was performed on live cells with a Zeiss Aviovert 10 inverted microscope (objective lens Zeiss Plan Neofluor 10/0.75).

Fig. 10 (A) illustrates the fluorescent flow cytometric analysis of EPO or EMP1 induced response in CHO-DUKX-B11 cells expressing EpoR-DHFR fragment fusions and labeled with fMTX. (A); upper panel; Cells transfected with EpoR-5aa-F[1,2] and -F[3]; middle panel, with EpoR-10aa-F[1,2] and -F[3] and lower panel, with EpoR-30aa-F[1,2] and -F[3]. Histograms are based on analysis of fluorescence intensity for 10,000 cells at flow rates of

approximately 1000 cells per second. Data were collected on a Coulter XL 4 color FACS analyzer (Coulter-Beckman) with stimulation with an argon laser tuned to 488 nm with emission recorded through a 525 nm bands filter. Histograms represent response in absence of ligands (black trace), with 10 nM Epo (dark gray trace) or 10 iM EMP1 (light gray trace). Preparation of cells for analysis was the same as described for microscopy (see Fig. 9), except that following the PBS wash, cells were gently trysinized, suspended in 500 lL of cold PBS supplemented with 10% FBS in order to increase cell viability and kept on ice prior to cytometric analysis within 20 minutes.

Fig. 10 (B) are the dose-response curves for Epo and EMP1 based on flow cytometric analysis of CHO DUKX-B11 cells expressing EpoR-5aa-F[1,2] and -F[3] as in (A), upper panel. Mean fluorescence intensity were determined for three separate samples at each ligand concentration (between 0.0003 nM and 100 nM, Epo (upper panel) or between 0.0003 iM and 100 lM for EMP1 (lower panel). X-axis is the mean fluorescence intensity relative to the maximum intensity observed and renormalized to zero for the minimum response. Traces through data points represents non-linear least-squares fit of results to a Langmuir isotherm determined in the computer program MacCurveFit (Kevin Raner Software) with a Quasi-Newton optimization routine (r^2 and residual error for Epo curve were 0.98 and 0.045, respectively and for EMP1 curve 0.99 and 0.022).

Fig. 11 is directed to fluorescence microscopy of COS-7 cells. Plasmids pMT3 harboring full length EpoR or EpoR or JAK2 fused *via* 5aa to F[1,2] or F[3] were created and COS-7 cells were transiently transfected or cotransfected with the different clones, and treated and analyzed as in figure 9.

Other objects, advantages and features of the present invention will become more apparent upon reading of the following non-restrictive description of preferred embodiments with reference to the accompanying drawings which are exemplary and should not be interpreted as limiting the scope of the present invention.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Selection of mDHFR for a PCA

In designing a protein-fragment complementation assay (PCA), it was sought to identify an enzyme for which the following is true: 1) An enzyme that is relatively small and monomeric, 2) for which structural and functional information exists, 3) for which simple assays exist for both *in vivo* and *in vitro* measurement, and 4) for which overexpression in eukaryotic and prokaryotic cells had been demonstrated. Murine DHFR (mDHFR) meets all of the criteria for a PCA listed above. Prokaryotic and eukaryotic DHFR is central to cellular one-carbon metabolism and is absolutely required for cell survival in both prokaryotes and eukaryotes. Specifically it catalyses the reduction of dihydrofolate to tetrahydrofolate for use in transfer of one-carbon units required for biosynthesis of serine, methionine, purines and thymidylate. The DHFRs are small (17 kD to 21 kD), monomeric proteins. The crystal structures of DHFR from various bacterial and eukaryotic sources are known and substrate binding sites and active site residues have been determined [111-114], allowing for rational design of protein fragments. The folding, catalysis, and kinetics of a number of DHFRs have been studied extensively[115-119]. The enzyme activity can be monitored *in vitro* by a simple spectrophotometric assay[120], or *in vivo* by cell survival in cells grown in the absence of DHFR end products. DHFR is specifically inhibited by the anti-folate drug trimethoprim. As mammalian DHFR has a 12000-fold lower affinity for trimethoprim than does bacterial DHFR[121], growth of bacteria expressing mDHFR in the presence of trimethoprim levels lethal to bacteria is an efficient means of selecting for reassembly of mDHFR fragments into active enzyme. High level expression of mDHFR has been demonstrated in transformed prokaryote or transfected eukaryotic cells[122-126].

Design Considerations

mDHFR shares high sequence identity with the human DHFR (hDHFR) sequence (91% identity) and is highly homologous to the *E. coli* enzyme (29% identity, 68% homology) and these sequences share visually superimposable tertiary structure[111]. Comparison of the crystal structures of mDHFR and hDHFR suggests that their active sites are essentially

identical[127,128]. DHFR has been described as being formed of three structural fragments forming two domains[129, 130] the adenine binding domain (residues 47 to 105 = fragment[2]) and a discontinuous domain (residues 1 to 46 = fragment[1] and 106 to 186 [3]; numbering according to the murine sequence). The folate binding pocket and the NADPH binding groove are formed mainly by residues belonging to fragments[1] and [2]. Fragment [3] is not directly implicated in catalysis.

Residues 101 to 108 of hDHFR, at the junction between fragment[2] and fragment[3], form a disordered loop which lies on the same face of the protein as both termini. It was chosen to cleave mDHFR between fragments [1,2] and [3], at residue 107, so as to cause minimal disruption of the active site and NADPH cofactor binding sites. The native *N*- terminus of mDHFR and the novel *N*-terminus created by cleavage occur on the same surface of the enzyme[112, 128] allowing for ease of *N*-terminal covalent attachment of each fragment to associating fragments such as the leucine zippers used in this study. Using this system, a leucine-zipper assisted assembly of the mDHFR fragments into active enzyme was obtained.

The present invention further illustrates that signaling by the Erythropoietin Receptor is mediated by a ligand-induced conformation change in constitutive receptor dimers. Erythropoietin and other cytokine receptors are thought to be activated through hormone-induced dimerization and autophosphorylation of JAK kinases associated with the receptor intracellular domains. Using an *in vivo* protein fragment complementation assay based on murine dihydrofolate reductase association with a fluorescent probe, applicants have discovered that constitutive erythropoietin receptor dimers exist in a conformation that prevents assocation of JAK2 but undergoes a ligand-induced conformation change that allows JAK2 to self-associate. These results are consistent with crystallographic evidence for the conformations of native and ligand-bound forms of the Erythropoietin receptor.

It is also known that Erythropoietin (Epo) regulates proliferation and differentiation of erythroid progenitors. Many disorders of erythroid proliferation are caused by genetic disorders of erythropoietin

biosynthesis or of genetic disruption of Epo synthesis or Epo receptor-mediated signal transduction (Foa, P., *Acta Haematologica* **86**, 162-8 (1991); Watowich, et al., *Annual Review of Cell & Developmental Biology* **12**, 91-128 (1996); Spivak, J.L., *Transactions of the American Clinical & Climatological Association* **102**, 232-42 (1990) and Lodish et al., *Cold Spring Harbor Symposia on Quantitative Biology* **60**, 93-104 (1995)). Such disorders include anemias due to renal failure, cancer chemotherapy and AZT treatment (Krantz, S.B., *Blood* **77**, 419-34 (1991)). The Epo receptor (EpoR) shares both structural and functional features with the cytokine receptor superfamily that includes the interleukins, human growth hormone (hGH) and colony stimulating factor CSF (D'Andrea et al., *Cell* **58**, 1023-1024 (1989); Bazan, J.F., *Proceedings of the National Academy of Sciences of the United States of America* **87**, 6934-8 (1990) and Stahl et al., *Cell* **74**, 587-590 (1993)). Functionally, the initial events in receptor-mediated signaling are the association, autophosphorylation and activation of one or two forms of the JAK family of tyrosine kinases (Chantler et al., *Biophysical Journal* **59**, 1242-50 (1991); Finbloom et al., *Cellular Signalling* **7**, 739-745 (1995); Ihle et al., *Annual Review of Immunology* **13**, 369-398 (1995); Witthuhn, W.A., et al., *Cell* **74**, 227-236 (1993)). Binding of JAKs to the receptors is mediated by common sequence elements (Box1 and Box2) of the intracellular domains of these receptors (Murikami, et al., *Proceedings of the National Academy of Science USA* **88**, 11349-1153 (1991) and Tanner et al., *Journal of Biological Chemistry* **270**, 6523-6530 (1995)). Crystal structures of hGH bound to GH receptor and EpoR bound to an agonist peptide EMP1 have shown that both the tertiary structures and oligomeric states of these two receptors are identical (Livnah, et al., *Science* **273**, 464-71 (1996) and De Vos, *Science* **255**, 306-312 (1992)). Both receptor-ligand complexes were found to be C2 symmetric homo-dimers that bound through two different surfaces of the receptors to one molecule of GH in the case of the GH receptor or a dimer of the EpoR agonist peptide. These studies, structures of other growth hormone receptors and biochemical analysis have led to the generally accepted dimerization model of growth factor-mediated receptor activation. Monomeric membrane-bound receptors remain inactive until ligand binds to and

oligomerizes the receptors. The activation event is autophosphorylation of intrinsic intracellular or non-covalently associated kinases brought into contact by the dimerization of receptors. Dimer- or oligomerization of receptors is a necessary but clearly not a sufficient condition for receptor activation. Other model receptors such as insulin and bacterial chemotactic Tar receptors exist as dimers in absence of ligand, and Tar receptors have been demonstrated to undergo ligand-induced change in conformation mechanically coupled to activation of the cytosolic kinase domain. Until now there was no direct biochemical or structural evidence that ligand-mediated activation of cytokine receptors could also involve an allosteric mechanism. Wilson et al., 1998 have solved the structure of unligated EpoR and shown that it is also a dimer, but with a dramatically different arrangement of the two subunits. Among the features of the unligated extracellular domain, is that the C-terminals of the monomers, the points of insertion into the membrane, are separated by 82 Å, compared to 34 Å in the ligated form. Assuming that these structures reflect the conformation of a constitutive dimer in cells it could be proposed that receptor activation by receptor would consist of a ligand-induced reorganization of the dimer that brings the intracellular domains into closer proximity and allows the associated JAK2s to come into contact and auto- phosphorylate (Fig. 8).

Applicants have also developed a fluorescent assay based on dimerization-induced complementation of designed fragments of the enzyme murine dihydrofolate reductase (DHFR) (Pelletier et al., *Protein Engineering* **10**, 89 (1997)) (Figure 8). The basis for the assay is that complementary fragments of DHFR when expressed and reassembled in cells, will bind to the high affinity (Kd= 100 pM) fluorescein-conjugated inhibitor methotrexate (fMTX) in a 1:1 complex. fMTX is retained in cells by this complex, while unbound is actively and rapidly transported out of the cells (Kaufman et al., *Journal of Biological Chemistry* **253**, 5852-60 (1978) and Israel et al., *Proceedings of the National Academy of Sciences of the United States of America* **90**, 4290-4 (1993)). In addition, binding of fMTX to DHFR results in an 4.5 fold increase in quantum yield. Bound fMTX and by inference reconstituted DHFR, can then be monitored by fluorescence microscopy, FACS or spectroscopy. Since the complex of fMTX

with DHFR is 1:1, measured fluorescence can be calibrated to det rmine average numbers of complexes in individual cells or averages in a population of cells. To test the allosteric model of receptor activation, it was reasoned that if the receptor transmembrane domain is separated by the distance observed in the crystal structure of unligated EpoR, then DHFR fragments fused to the C-terminal of the transmembrane domains will complement only if ligand induces the necessary conformation change that allows the fragments to come into contact. Furthermore, the absolute regio- and stereospecific requirement that fragments be sufficiently close to fold-reassemble into the enzyme is three dimensional structure means that a false response that might occur if fused is unlikely (interacting proteins are merely proximal). In addition, insertion of flexible linker peptides of a critical length between the transmembrane domain and the fragments should result in constitutive complementation, insensitive to ligand. Based on the EpoR crystal structure, the minimum length of linker necessary for a constitutive response would be 10 amino acids, assuming the length of an average peptide bond is ~4 Å and the distance separating the fragments is 82 Å. Longer linkers should result in complementation, independent of ligand. Linkers of 5, 10 and 30 amino acids corresponding to extended lengths of 20, 40, and 120 Å, respectively were thus used (Fig. 8).

The present invention is illustrated in further detail by the following non-limiting examples.

EXAMPLE 1

EXPERIMENTAL PROTOCOL

DNA Constructs

Mutagenic and sequencing oligonucleotides were purchased from Gibco BRL. Restriction endonucleases and DNA modifying enzymes were from Pharmacia and New England Biolabs. The mDHFR fragments carrying their own iN-frame stop codon were subcloned into pQE-32 (Qiagen), downstream from and iN-frame with the hexahistidine peptide and a GCN4 leucine zipper (Fig 1; Fig. 2). All final constructs were based on the Qiagen pQE series of vectors, which contain an inducible promoter-operator element (tac),

a consensus ribosomal binding site, initiator codon and nucleotides coding for a hexahistidine peptide. Full-length mDHFR is expressed from pQE-16 (Qiagen).

Expression vector harboring the GCN4 leucine zipper

Residues 235 to 281 of the GCN4 leucine zipper (a SalI/BamHI 254 bp fragment) were obtained from a yeast expression plasmid pRS316[9]. The recessed terminus at the BamHI site was filled-in with Klenow polymerase and the fragment was ligated to pQE-32 linearized with SalI/HindIII(filled-in). The product, construct Z, carries an open reading frame coding for the sequence Met-Arg-Gly-Ser followed by a hexahistidine tag and 13 residues preceding the GCN4 leucine zipper residues.

Creation of DHFR fragments

The eukaryotic transient expression vector, pMT3 (derived from pMT2)[16], was used as a template for PCR-generation of mDHFR containing the features allowing subcloning and separate expression of fragment[1,2] and fragment[3]. The megaprimer method of PCR mutagenesis[29] was used to generate a full-length 590 bp product. Oligonucleotides complementary to the nucleotide sequence coding for the *N*- and *C*-termini of mDHFR and containing a novel BspEI site outside the coding sequence were used as well as an oligonucleotide used to create a novel stop codon after fragment[1,2], followed by a novel SpeI site for use in subcloning fragment[3].

Construction of a new multiple cloning region and subcloning of DHFR fragments [1,2] and [3]

Complementary oligonucleotides containing the novel restriction sites: SnaBI, NheI, SpeI and BspEI, were hybridized together resulting in 5' and 3' overhangs complementary to EcoRI, and inserted into pMT3 at a unique EcoRI site. The 590 bp PCR product (described above) was digested with BspEI and inserted into pMT3 linearized at BspEI, yielding construct [1,2,3]. The 610 bp BspEI/EcoNI fragment (coding for DHFR fragment[1,2], followed by a novel stop and fragment[3] up to EcoNI) was filled in at EcoNI and subcloned

into pMT3 opened with BspEI/HpaI, yielding construct F[1,2]. The 250 bp SpeI/BspEI fragment of construct [1,2,3] coding for DHFR fragment[3] (with no in-frame stop codon) was subcloned into pMT3 opened with the same enzymes. The stop codon of the wild-type DHFR sequence, downstream from fragment[3] in pMT3, was inserted as follows. Cleavage with EcoNI, present in both the inserted fragment[3] and the wild-type fragment[3], removal of the 683 bp intervening sequence and religation of the vector yielded a construct of fragment[3] with the wild-type stop codon, construct F[3].

Creation of the expression constructs

The 1051 bp and the 958 bp SnaBI/XbaI fragments of constructs F[1,2] and F[3], respectively, were subcloned into construct Z opened with BglII(filled-in)/NheI, yielding constructs Z-F[1,2] and Z-F[3] (Fig. 2). For the Control expression construct, the 180 bp XmaI/BspEI fragment coding for the zipper was removed from construct Z-F[1,2], yielding construct Control-F[1,2] (Fig. 2).

Creation of Stability Mutants

Site-directed mutagenesis was performed[30] to produce mutants at Ile114 (numbering of the wild-type mDHFR). The mutagenesis reaction was carried out on the KpnI/BamHI fragment of construct Z-F[3] subcloned into pBluescript SK+ (Stratagene), using oligonucleotides that encode a silent mutation producing a novel BamHI site. The 206 bp NheI/EcoNI fragment of putative mutants identified by restriction was subcloned back into Z-F[3]. The mutations were confirmed by DNA sequencing.

***E. coli* Survival Assay**

E. coli strain BL21 carrying plasmid pRep4 (from Qiagen, for constitutive expression of the lac repressor) were made competent, transformed with the appropriate DNA constructs and washed twice with minimal medium before plating on minimal medium plates containing 50 mg/ml kanamycin, 100 mg/ml ampicillin and 0.5 mg/ml trimethoprim. One half of each transformation

mixture was plated in the absence, and the second half in the presence, of 1 mM IPTG. All plates were placed at 37°C for 66 hrs.

***E. coli* Growth Curves**

Colonies obtained from cotransformation were propagated and used to inoculate 10 ml of minimal medium supplemented with ampicillin, kanamycin as well as IPTG (1mM) and trimethoprim (1 µg/µl) where indicated. Cotransformants of Z-F[1,2] + Z-F[3:Ile114Gly] were obtained under non-selective conditions by plating the transformation mixture on L-agar (+ kanamycin and ampicillin) and screening for the presence of the two constructs by restriction analysis. All growth curves were performed in triplicate. Aliquots were withdrawn periodically for measurement of optical density. Doubling time was calculated for early logarithmic growth (OD 600 between 0.02 and 0.2).

Protein Overexpression and Purification

Bacteria were propagated in Terrific Broth[31] in the presence of the appropriate antibiotics to an OD600 of approximately 1.0. Expression was induced by addition of 1 mM IPTG and further incubation for 3 hrs. For analysis of crude extract, pellets from 150 ml of induced cells were lysed by boiling in loading dye. The lysates were clarified by microcentrifugation and analyzed by SDS-PAGE32. For protein purification, a cell pellet from 50 ml of induced *E. coli* cotransformed with constructs Z-F[1,2] and Z-F[3] was lysed by sonication, and a denaturing purification of the insoluble pellet undertaken using Ni-NTA (Qiagen) as described by the manufacturer. The proteins were eluted with a stepwise imidazole gradient. The fractions were analyzed by SDS-PAGE.

RESULTS

Design of mDHFR fragments for a PCA

mDHFR shares high sequence identity with the human DHFR (hDHFR) sequence. As the coordinates of the murine crystal structure were not available, the design considerations were based on the hDHFR structure. DHFR has been described as comprising three structural fragments forming two domains: the adenine binding domain (F[2]) and a discontinuous domain (F[1] and F[3])[13,18]. The folate binding pocket and the NADPH binding groove are formed mainly by residues belonging to F[1] and F[2]. Residues 101 to 108 of

hDHFR form a disordered loop which lies on the same face of the protein as both termini. This loop occurs at the junction between F[2] and F[3]. By cleaving mDHFR at residue 107, F[1,2] and F[3] were created, thus causing minimal disruption of the active site and substrate binding sites. The native *N*-terminus of mDHFR and the novel *N*-terminus created by cleavage were covalently attached to the *C*-termini of GCN4 leucine zippers (Fig. 1).

***E. coli* Survival Assays**

Figure 2 illustrates the general features of the expressed constructs and the nomenclature used in this study. Figure 3 (panel A) illustrates the results of cotransformation of bacteria with constructs coding for Z-F[1,2] and Z-F[3], in the presence of trimethoprim, clearly showing that colony growth under selective pressure is possible only in cells expressing both fragments of mDHFR. There is no growth in the presence of either Z-F[1,2] or Z-F[3] alone. Induction of protein expression with IPTG is essential for colony growth (Fig. 3A). The presence of the leucine zipper on both fragments of mDHFR is essential as illustrated by cotransformation of bacteria with both vectors coding for mDHFR fragments, only one of which carries a leucine zipper (Fig. 3A). It should be noted that growth of control *E. coli* transformed with the full-length mDHFR is possible in the absence of IPTG due to low levels of expression in uninduced cells.

Confirmation of the presence of both plasmids in bacteria, now able to grow with trimethoprim, was obtained from restriction analysis of the plasmid DNA purified from isolated colonies. Figure 4 (A) reveals the presence of the 1200 bp HincII restriction fragment from construct Z-F[1,2] as well as the 487 and 599 bp HincII restriction fragments from construct Z-F[3]. Also present is the 935 bp HincII fragment of pRep4. Overexpression of the fusion proteins is illustrated in Figure 4 (B). In all cases, overexpression of a protein of the expected molecular weight is apparent on SDS-PAGE of the crude lysate. Purification of the coexpressed proteins under denaturing conditions yielded two bands of apparent homogeneity upon analysis by Coomassie-stained SDS-PAGE (Fig. 4B).

Stability Mutants

Applicants generated mutants of F[3] to test whether reconstitution of mDHFR activity by fragment assembly was specific. Protein stability can be reduced by changing the side-chain volume in the hydrophobic core of a protein[9], [22-25]. Residue Ile114 of mDHFR occurs in a core β-strand at the interface between F[1,2] and F[3], isolated from the active site. Ile 114 is in van der Waals contact with Ile51 and Leu93 in F[1,2][11]. Ile 114 was mutated to Val, Ala, or Gly. Figure 3 (panel B) illustrates the results of cotransformation of *E. coli* with construct Z-F[1,2] and the mutated Z-F[3] constructs. The colonies obtained from cotransformation with Z-F[3:Ile114Ala] grew more slowly than those cotransformed with Z-F[3] or Z-F[3:Ile114Val] (see inset to Fig. 3B). No colony growth was detected in cells cotransformed with Z-F[3:Ile114Gly]. The number of transformants obtained was not significantly different in the case where colonies were observed, implying that cells cotransformed with Z-F[1,2] and either Z-F[3], Z-F[3:Ile114Val] or Z-F[3:Ile114Ala] have an equal survival rate. Overexpression of the mutants Z-F[3:Ile114X] was in the same range as Z-F[3], as determined by Coomassie-stained SDS-PAGE (data not shown).

The relative efficiency of reassembly of mDHFR fragments was also compared by measuring the doubling time of the cotransformants in liquid medium. Doubling time in minimal medium was constant for all transformants (data not shown). Selective pressure by trimethoprim in the absence of IPTG prevented growth of *E. coli* except when transformed with pQE-16 coding for full-length DHFR due to low levels of expression in uninduced cells. Induction of mDHFR fragment expression with IPTG allowed survival of cotransformed cells (except in the case of Z-F[1,2] + Z-F[3:Ile114Gly], although the doubling times were significantly increased relative to growth in the absence of trimethoprim. The doubling time measured for cells expressing Z-F[1,2] + Z-F[3], Z-F[1,2] + Z-F[3:Ile114Val] and Z-F[1,2] + Z-F[3:Ile114Ala] were 1.6-fold, 1.9-fold and 4.1-fold, higher respectively, than the doubling time of *E. coli* expressing pQE-16 in the absence of trimethoprim and IPTG. The presence of IPTG unexpectedly prevented growth of *E. coli* transformed with full-length mDHFR. Growth was partially restored by addition of the folate metabolism end-

products thymine, adenine, pantothenate, glycine and methionine (data not shown). This suggests that induced overexpression of mDHFR was lethal to *E. coli* when grown in minimal medium as a result of depletion of the folate pool by binding to the enzyme.

In another embodiment, applicants make point mutations in the GCN4 leucine zipper of Z-F[1,2] and Z-F[3], for which direct equilibrium and kinetic parameters are known and correlating these known values with parameters derived from the PCA (Pelletier and Michnick, in preparation). Comparison of cell growth rates in this model system with rates for a DHFR PCA using unknowns would give an estimate of the strength of the unknown interaction. This should enable the determination of estimates of equilibrium and kinetic parameters for a specific protein-protein interaction.

The present invention has illustrated and demonstrated a protein-fragment complementation assay (PCA) based on mDHFR, where a leucine zipper directs the reconstitution of DHFR activity. Activity was detected by an *E. coli* survival assay which is both practical and inexpensive. This system illustrates the use of mDHFR fragment complementation in the detection of leucine zipper dimerization and could be applied to the detection of unknown, specific protein-protein interactions *in vivo*.

<u>*E. coli* Aminoglycoside kinase: Optimization and Design of a PCA using an Exonuclease-Molecular Evolution Strategy</u>

Although applicants have demonstrated that the engineering/design strategy described above can be used to produce complementary enzyme fragments, it is obvious that proteins did not evolve in such a way that such fragments would be expected to have optimal physical characteristics, including solubility, foldability (fast folding), protease resistance, or enzymatic activity. An alternative embodiment to the engineering/design strategy is the endonuclease/evolution approach. This strategy can be used by itself or in conjunction with the engineering/design strategy. The advantages of this approach are that in principle, prior knowledge of the protein structure is not necessary, that the optimal fragments are chosen for PCA and that these fragments will also have optimal characteristics. Following selection of optimal

complementary fragments, the fragments are exposed to multiple rounds of random mutagenesis. Mutagenesis is achieved by suboptimal PCR combined with chemical mutagenesis or DNA shuffling (Stemmer, W. P. C. 1994, Proc, Natl, Acad, Sci. USA 91:10747-10751). The overall strategy is described for the case of aminoglycoside kinase (AK), an example of antibiotic resistance marker that can be used for dominant selection of prokaryotic cells such as *E. coli* or eukaryotic cells such as yeast or mammalian cell lines. The structure of AK is already known, and so strategy (1) would be possible, however a combination of strategy (1) as defined for DHFR above, in conjunction with strategy (2) was chosen.

EXPERIMENTAL PROTOCOL

The optimization/selection procedure is as follows.

Generation of of library of AK fragments based on products of Exonuclease digestion

Nested sets of deletions are created at the 5' and the 3' ends of the AK gene. In order to create unidirectional deletions, unique restriction sites are introduced in the regions flanking the AK gene. At the 5' and 3' termini, an "outer" sticky site with a protruding 3' terminus (Sph I and Kpn I, respectively) and an "inner" sticky site with recessed 3' terminus (Bgl II and Sal I, respectively) are added by PCR. Cleavage at Sph I and Bgl II (or Kpn I and Sal I) results in creation of a protruding terminus leading back to the flanking sequence and a recessed terminus leading into the AK gene. Digestion with *E. coli* exonuclease III and S1 nuclease (Henikoff, S. 1987, Methods in Enzymology 155:156-165) yields a set of nested deletions from the recessed terminus only. Thus, 10 mg of DNA is digested with Sph I and Bgl II (or Kpn I and Sal I), phenol-chloroform extracted, and 12.5 U exonuclease III added. At 30 sec intervals over 10 min, aliquots are taken and put into solution with 2 U S1 nuclease. The newly created ends are filled in with T4 DNA polymerase (0.1 U per sample) and the set of vectors closed back by blunt-ended ligation (10 U ligase per sample). The average length of the deletion at each time point is determined by restriction analysis of the sets. This yields sets of AK genes d leted from the 5' or the 3' termini. This manipulation is undertaken directly in

the pQE-32-Zipper constructs, such that the products can be used directly in activity screening.

Screening for AK activity

As a first step in determining the requirements for fragment complementation, the minimum *N*-terminal and *C*-terminal fragments of AK that, alone, are active must be determined. Sets of deletions are individually transformed into *E. coli* BL21 cells and expression of the AK fragments is induced by IPTG. The sets where a significant number of colonies appear in the presence of G418 serve to indicate the approximate length of *N*- and *C*-terminal AK fragments which retain activity. Fragment complementation must therefore be undertaken with fragments taken from within these limits. The zipper-directed fragment complementation is detected as follows: appropriate sets of deletions, or pools of sets, are cotransformed into BL21, expression is induced with IPTG and growth in the presence of varying G418 concentrations is monitored. Large colonies which grow in the presence of high G418 concentrations are selected as giving the most efficiently complementing products.

Directed evolution of optimal AK fragments using "DNA shuffling

After optimal fragments have been selected, the individual fragments are removed by restriction digestion at Sph I and Kpn I allowing for 5' and 3' constant priming regions flanking the *N*- or *C*-terminal complementary fragments of AK. These oligonucleotides (2-4 µg) are digested with DNaseI (0.005 units/ul, 100 ul) and fragments of 10-50 nucleotides are extracted from low melting point agarose. PCR is then performed with the fragmented DNA, using Taq polymerase (2.5 units/ul) in a PCR mixture containing 0.2 mM dNTPs, 2.2 mM Mg_2Cl (or 0 mM for subuptimal PCR), 50 mM KCl, 10 mM Tris.HCl, pH 9.0, 0.1% TritonX-100. A PCR program of 94°C/60 sec.; 94°C 30 sec.; 55°C 30 sec.; 72°C 30 sec. times 30 to 50; 72°C 5 min. Samples are taken every 5 cycles after 25 cycles to monitor the appearance of reassembled complete fragments on agarose gel. The primerless PCR product is then diluted 1:40 or 1:60 and used as template for PCR with 5', 3' complementary constant region oligos as primers for a further 20 cycles. Final product is

restriction digested with Sph I and Kpn I and the products subcloned back into pQE32-Zipper to yield the final library of expression plasmids. As before, *E. coli* BL21 cells are sequentially transformed with *C*-terminal or *N*-terminal complementary fragment-expression vectors at an estimated efficiency of 10^9 and finally cells cotransformed with the complementary fragment. *E. coli* are grown on agarose plates containing 1 µg/ ml G418 and after 16 hours the largest colonies are selected and grown in liquid medium at increasing concentrations of G418. Those clones showing the maximal resistance to G418 are then selected and if maximum resistance or greater is reached the evolution is terminated. Otherwise the DNA shuffling proceedure is repeated. Finally, optimal fragments are sequenced and physical properties and enzymatic activity are assessed. This optimized AK PCA is now ready to test for dominant selection in any other cell type including yeast and mammalian cell lines. This strategy can be used to develop any PCA based on enzymes that impart dominant or recessive selection to a drug or toxin or to enzymes that produce a colored or fluorescent product. In the later two cases the end point of the evolution process is at minimum, reatainment of signal for the intact, wild type enzyme or enhancement of the signal. This strategy can also be used in the absense of knowledge of the enzyme structure, whether the enzyme has a mono-, di- or multimeric structure. However, knowledge of the enzyme structure does not preclude applying this strategy as well, as described below.

As can be appreciated, knowledge of the enzyme structure can provide a more efficient way of using molecular evolution to design a PCA. In this case, the enzyme structure is used to define minimal domains of the protein in question, as was done for DHFR. Instead of generating fragments of completely random length for the *N*- or *C*-terminal fragments, fragments that at a minimum will code for one of the two domains are selected during the exonuclease phase. For instance, in the case of AK, two well defined domains can be discerned in the structure consisting of residues 1-94 in the *N*-terminus and residues 95-267 in the *C*-terminus. Endonuclease digestions are performed as above, but reaction products are selected that will minimally code

for one of the two domains. These are then the starting points for fragment selection and evolution cycles as described above.

Heteromeric Enzyme PCA

A further embodiment of the invention relates to PCA based on using heterodimeric or heteromultimeric enzymes in which the entire catalytic machinery is contained within one independently folding subunit and the other subunit provides stability and/or a cofactor to the enzymatic subunit. In this embodiment of PCA, the regulatory subunit is split into complementary fragments and fused to interacting proteins. These fragments are co-transformed/transfected into cells along with the enzyme subunit. As with single enzyme PCA, described for DHFR and AK, reconstitution and detection of enzyme activity is dependent on oligomerization domain-assisted reassembly of the regulatory subunit into its native topology. However, the reconstituted subunit then interacts with the intact enzymatic subunit to produce activity. This approach is reminiscent of the USPS system, except that it has the advantage that, in this case, the enzyme is not a constitutive cellular enzyme, but an exogenous gene product. As such there is no problem with background activity from the host cell, the enzyme can be expressed at higher levels than a natural gene and can also be modified to be directed to specific subcellular compartments (by subcloning compartment-specific signal peptides onto the *N*- or *C*-termini of the enzyme and subunit fragments). The specific advantage of this approach is that while the single enzyme strategy may lead to suboptimal enzymatic activity, in this approach, the enzyme folds independently and may in fact act as a chaperone to the fragmented regulatory subunit, aiding in its refolding. In addition, folding of the fragments may not need be complete in order to impart regulation of the enzyme. This approach is realized by a colorimetric/fluorometric assay that was developed and based on the *Streptomyces* tyrosinase. This enzyme catalyzes the conversion of tyrosine to deoxyphenylalanine (DOPA). The reaction can be measured by conversion of fluorocinyl-tyrosine to the DOPA form. The active enzyme consists of two subunits, the catalytic domain (Melc2) and a copper binding domain (Melc1). Melc1 is a small protein of 14 kD that is absolutely required for Melc2 activity.

In one assay which was developed, the Melc1 protein is split into two fragments that serve as the complementation part of the PCA. These fragments, fused to oligomerization domains, are coexpressed with Melc2, and the basis of the assay is that Melc2 activity is dependent on complementation of the Melc1 fragments. Stoichiometries of protein complexes can also be addressed (i.e. whether a complex consists of two or three proteins) as follows. One fuses two proteins to the two Melc1 fragments and a third to intact Melc2. It thus can be shown that the minimum complementary active complex of the tyrosinase will require that all three components and therefore a trimer is necessary. A key aspect of this approach is that specific interactions can easily be demonstrated by making one component, specifically the protein-Melc2 fusions' catalytic subunit, dependent on the other components, by underexpressing it in the background of overexpressed Melc1 fragment-protein fusions.

Multimer Disruption-Based PCA

Although applicants have described only fragment complementation of intact proteins, protein domains or subunits as comprising PCA, an alternate embodiment relates to PCAs based on the disruption of the interface between, for instance a dimeric enzyme that requires stable association of the subunits for catalytic activity. In such cases, selective or random mutagenesis at the subunit interface would disrupt the interaction and the basis of the assay would be that oligomerization domains fused to the subunits would provide the nessesary binding energy to bring the subunits together into a functional enzyme.

Vector Design in Application to PCAs

The PCA strategies listed thus far have used two-plasmid transformation strategies for expression of complementary fragments. This approach has some advantages, such as using different drug resistance markers to select for optimal incorporation of genes, for instance, in transformed or transfected cells or for optimum transformation of complementary plasmids into bacteria and control of expression levels of PCA fragments using different promoters. However, single plasmid strategies have advantages in terms of simplicity of transfection/ transformation. Protein expression levels can be

controlled in different ways, while drug selection can be achieved in one of two ways: In the case of PCAs based on survival assay using enzymes that are drug resistance markers themselves, such as AK, or where the enzyme complements a metabolic pathway, such as DHFR, no additional drug resistance gene need be incorporated in the expression plasmids. If however the PCA is based on an enzyme that produces a colored or fluorescent product, such as tyrosinase or firefly luciferase, an additional drug resistance gene must be expressed from the plasmid. Expression of PCA complementary fragments and fused cDNA libraries/target genes can be assembled on single plasmids as individual operons under the control of separate inducible or constitutive promotors, or can be expressed polycistronically. In *E. coli*, polycistronic expression can be achieved using known intercoding region sequences. For instance, the region in the mel operon from which the tyrosinase melc1-melc2 genes were derived can be used. Indeed, this region was shown to be expressed at high levels in *E. coli* under the control of a strong (tac) promoter. Genes could also be expressed and induced by independent promoters, such as tac and arabinose.

For mammalian expression systems, single plasmid systems can be used for both transient or stable cell line expression and for constitutive or inducible expression. Further, differential control of the expression of one of the complementary fragment fusions, usually the bait-fused fragment, can be controlled to minimize expression. This will be important in reducing background non-specific interactions. Examples of differential control of complementary fragment expression include the following strategies:

i) In polycistronic expression, transient or stable, expression of the second gene will necessarily be less efficient. This in itself could thus serve to limit the quantity of one of the complementary fragments. Alternatively, the expression of the first gene product can be limited by mutation of an upstream donor/splice site, while the second gene can be put under the control of a retroviral internal initiation site, such as that of ECMV of poliovirus, to enhance expression.

ii) Individual complementary fragment-fusion pairs can also be put under the control of commercially available inducible promoters, including for example those based on Tet-responsive PhCMV*-1 promoter, and/or steroid receptor

response elements. In such a system the two complementary fragment genes can be turned on and expression levels controlled by dose dependent expression with the inducer (tetracycline and steroid hormones, respectively).

EXAMPLE 2

Applications of the PCA strategy to detect novel gene products in biochemical Pathways and to map such pathways

Among the greatest advantage of PCA over other molecular interaction screening methods is that they are designed to be versatile enough to enable a performance both *in vivo* and in any type of cell. This feature is crucial if the goal of applying a technique is to identify novel interactions from libraries and simultaneously be able to determine if the interactions observed are biologically relevant. The detailed example given below, and other examples at the end of this section illustrate how it is that validation of interactions with PCA is possible. In essence, this is achieved as follows. In biochemical pathways, such as hormone receptor-mediated signaling, a cascade of enzyme-mediated chemical reactions are triggered by some molecular event (i.e. hormone binding to its membrane surface receptor). Enzyme interactions with protein substrates and protein-protein or protein-nucleic acid interactions with enzyme-modified substrates then occurs. Such biochemical signaling cascades generally only occur in specific cell types and model cell lines for studying these processes. Therefore, to detect induced interactions, such as with known proteins in a pathway with yet unidentified proteins, one obviously needs to perform such screening in appropriate model cell lines and in the correct cellular compartment. Only the PCA strategy can be used in a general way to do this. Protein-molecular interaction techniques such as yeast two- or three-hybrid techniques cannot be performed in a context where such events occur, except in the limiting case of nuclear interaction in yeast or interactions that are not triggered. There do exist mammalian two-hybrid techniques making it possible to detect induced protein interactions, but only if the proteins involved can be simultaneously activated, transported to the nucleus and interact with their partners. PCAs do not have these limitations since they do not require additional cellular machinery available only in specific compartments. A further point is that by performing the PCA strategy in appropriate model cell types, it is also possible to introduce appropriate positive and negative controls for studying a particular pathway. For instance, for a hormone signaling pathway, it is likely that hormone signaling

agonists and antagonists or dominant-negative mutants of signaling cascade proteins that are upstream or act in parallel to the events being examined in the PCA would be known. These reagents could be used to determine if novel interactions detected by the PCA are biologically relevant. In general then, interactions that are detected only if a hormone is introduced, but are not seen if an antagonist is simultaneously introduced could be hypothesized to represent interactions relevant to the process under study.

Below is a detailed description of an application of the DHFR that illustrates these points, as well as further examples where the PCA strategy could be used.

Application of the DHFR PCA to Mapping Growth Factor-Mediated Signal Transduction Pathways

One of the earliest detectable events in growth factor-activated cell proliferation is the serine phosphorylation of the S6 protein of the 40S ribosomal subunit. The discovery of serine/threonine kinases that specifically phosphorylate S6 have considerably aided in identifying novel mitogen mediated signal transduction pathways. The serine/threonine kinase p70S6k has been identified as a specific S6 phosphorylase[131-136]. p70S6k is activated by serine and threonine phosphorylation at specific sites in response to several mitogenic signals including serum in serum starved cells, growth factors including insulin and PDGF, and mitogens such as phorbol esters. Considerable effort has been made over the last five years to determine how p70/p85S6k are activated in response to mitogens. Two receptor-mediated pathways have been implicated in p70S6k activation, one associated with the phosphatidylinositol-3-kinase (PI(3)k) and the other with the PI(3)k homologue mTOR[137-144]. Key to understanding of this proposal, is the fact that the role of these enzymes in activation of p70S6k was determined by effects of two natural products on phosphorylation and enzyme activity: rapamycin, which indirectly inhibits mTOR activity, and wortmannin, which directly inhibits PI(3)k activity. It is also important to note that no direct upstream kinases or other regulatory proteins of p70S6k have been identified to this date.

The interactions of p70S6k with its known substrat S6 can be studied as a test system for the DHFR PCA in *E. coli* and in mammalian cell lines. One can also seek to identify novel interactions with this enzyme that would lead to new insights into how this important enzyme is regulated. Also, since activation of the enzyme is mediated by multiple pathways that can be selectively inhibited with specific drugs, this is an ideal system to test PCAs as methods to distinguish induced versus constitutive protein-protein interactions.

a) Testing of the *E. coli* survival assay: Interaction of p70S6k with S6

This test is ideal, because the apparent Km (= 250 nM) of p70S6k for S6 protein[145] is approximately the same as the Kd for leucine zipper-forming peptides from GCN4 used in the test system herein disclosed. However, a constitutively active form of the enzyme was used for the following tests. An *N*-terminal truncated form of the enzyme D77-p70S6k, is constitutively active and will be used in these studies147.

Methodology: D77-p70S6k-F[1,2] fusion and D77-p70S6k-F[3] fusion, or F[1,2] and D77-p70S6k-F[3] fusion (as a control) will be cotransformed into *E. coli* and the cells grown in minimal medium in the presence of trimethoprim. Colonies will be selected and expanded for analysis of kinase activity against 40S ribosomal subunits, and for coexpression of the two proteins.

b) Modification of the bacterial survival assay for library screening: Identification of Novel Interacting Proteins

Screening an expression library for interactions with a given target (p70S6k-D77, in this case) will be straightforward in this system, given that the only steps involved are: 1-construction of the fusion-expression library as a fusion with mDHFR fragment[3]; 2-transformation of the library in *E. coli* BL21 harboring pRep4 (for constitutive expression of the lac repressor; this is required in the case of a protein product which is toxic to the cells) and a plasmid coding for the fusion: p70S6k-D77-[1,2]; 3-plating on minimal medium in the presence of trimethoprim and IPTG; 4-selection of any colonies that grow, propagation and isolation of plasmid DNA, followed by sequencing of DNA

inserts; 5-purification of unknown fusion products via the hexaHis-tag and sizing on SDS-PAGE.

Methodology:

The overall strategy is illustrated in Figure 5. 1-Construction of a directional fusion-expression library: i-cDNA production: One can isolate poly(A)+ RNA from BA/F3 cells (B-lymphoid cells) because these cells have successfully been used in the study of the rapamycin-sensitive p70S6k activation cascade[139]. To enrich for full-length mRNA, the mRNA will be affinity-purified via the 5' cap structure by the CAPture method[148]. Reverse transcription will be primed by a "Linker Primer": it has a poly(T) tail to prime from the poly(A) mRNA tail, and an XhoI site for later use in directional subcloning of the fragments. The first strand is then methylated. After second strand synthesis and blunting of the products, "EcoRI Adapters" are added. Digestion of the linkers with EcoRI and XhoI (the inserts are protected by methylation) produces full-length cDNA ready for directional insertion in a vector opened with EcoRI and XhoI. Since the success of library screening depends largely on the quality of the cDNA produced, the above methods will be used as they have proven to consistently produce high-quality cDNA libraries. ii-Insertion of the cDNA into vectors: The library will be constructed as a C-terminal fusion to mDHFR F[3] in vector pQE-32 (Qiagen), as high levels of expression of mDHFR fusions were obtained from this vector in BL21 cells. Three such vectors will be created, differing at their 3' end, which is the novel polycloning site that was engineered (described earlier, under Methods), carrying either 0, 1, or 2 additional nucleotides. This allows read-through from F[3] into the library fragments in all 3 translational reading frames. The cDNA fragments will be directionally inserted at the EcoRI and XhoI sites in all three vectors at once. 2, 3, 4, and 5-These steps have been described earlier, under Results, apart from the final sequencing of clones identified using sequencing primers specific to vector sequences flanking sites of library insertion. The protein purification will also be as described earlier, by a one-step purification on Ni-NTA (Qiagen). If the product size is more than 15 kDa over the molecular weight of the DHFR

component (equal to a cDNA insert of more than 450bp), the inserts will have to be sequenced (i.e. at the Sheldon Biotechnology Center, McGill University).

c) Development of the Eukaryotic Assay

The transformation of the system described above, is useful to produce an equivalent assay for use in eukaryotic cells. The basic principle of the assay is the same: the fragments of mDHFR are fused to associating domains, and domain association is detected by reconstitution of DHFR activity in eukaryotic cells (Figure 5).

Creation of the expression constructs

The DNA fragments coding for the GCN4-zipper-mDHFR fragment fusions were inserted as one piece into pMT3, a eukaryotic transient expression vector[126]. Expression of the fusion proteins in COS cells was apparent on SDS-PAGE after 35[S]Met labeling.

Survival assays in eukaryotic cells

Two systems can be used for detection of mDHFR reassembly, in parallel: i- CHO-DUKX B11 cells (Chinese Hamster Ovary cell line deficient in DHFR activity) are cotransfected with GCN4-zipper-mDHFR fragment fusions. The cells are grown in the absence of nucleotides; only cells carrying reconstituted DHFR will undergo normal cell division and colony formation. ii- Methotrexate (MTX)-resistant mutants of mDHFR have been created, with the goal of transfecting cells that have constitutive DHFR activity such as COS and 293 cells. F[1,2] were mutated in order to incorporate, one at a time, each of five mutations that significantly increase the Ki (MTX): Gly15Trp, Leu22Phe, Leu22Arg, Phe31Ser and Phe34Ser (numbering according to the wild-type mDHFR sequence). These mutations occur at varying positions relative to the active site and relative to F[3], and have varying effects on Km (DHF), Km (NADPH) and Vmax of the full-length mammalian enzymes in which they were. Mutants Z-F[1,2: Leu22Phe], Z-F[1,2: Leu22Arg] and Z-F[1,2: Phe31Ser] all allowed for bacterial survival with high growth rates when cotransformed with Z-F[3] (results not shown).The five mutants will be tested in eukaryotic cells, in reconstitution of mDHFR fragments to produc enzyme that can sustain COS or 293 cell growth while under the selective

pressure of MTX, which will eliminate background due to activity of the native enzyme. The mutations offers an advantage in selection while presenting no apparent disadvantage with respect to reassembly of active enzyme. If the reconstituted mDHFR produced in either of the survival assays allows eukaryotic cell growth that is significantly slower than growth with the wild-type enzyme, thymidylate will be added to the growth medium to partially relieve the selective pressure offered by the lack of nucleotides.

d) Testing of the eukaryotic survival assay

It is necessary at the outset to test whether induced interactions with p70S6k can be detected. One can use the same test system as that for the *E. coli* test system described above: Induction of association of p70S6k with S6 protein.

Methodology:

mDHFR Leu22Phe mutant S6-F[1,2] and p70S6k-F[3], or F[1,2] and p70S6k-F[3] (as a control) will be cotransfected into COS cells and the cells will be serum starved for 48 hours followed by replating of cells at low density in serum and MTX. Colonies will be selected and expanded for analysis of kinase activity against 40S ribosomal subunits, and for coexpression of the two proteins. Further controls will be performed for inhibition of protein association with wortmannin and rapamycin.

e) Modification of the eukaryotic survival assay for library screening

An important part of the work required in creating a library for use in eukaryotic cells will have been accomplished already, as the EcoRi/XhoI directional cDNA produced by the Stratagene "cDNA Synthesis Kit" can directly be inserted directionally into the Stratagene Zap Express vector.

Methodology:

Steps 1 through 5 are parallel to those for the bacterial library screening (above). 1-Again, the library is constructed as a C-terminal fusion to mDHFR F[3]. F[3] (with no stop codon) will be inserted in frame in Zap Express, followed by insertion of the novel polylinkers allowing expression of the inserts in all three reading frames (described above), and by the EcoRI/XhoI directional cDNA. This bacteriophage library will be propagated and treated with the

Stratagene helper phage to excise a eukaryotic expression phagemid vector (pBK-CMV) carrying the fusion inserts. 2-Cotransfection of the library and p70S6k-F[1,2] constructs in eukaryotic cells: the screening in COS or 293 cells was performed, as these are responsive to serum in activating the p70S6k signaling pathway. Selection experiments will be performed as described for the S6 test system above. 3-Propagation, isolation and sequencing of the insert DNA will be undertaken. 4-The cloned fusion proteins will be sized on SDS-PAGE by direct visualization after 35S-Met/Cys labeling, or by Western blotting using a commercial polyclonal antibody to mDHFR.

Generalization of the Strategy

The scheme for detecting partners for the protein p70S6k can be applied to studies of any biochemical pathway in any living organism. Such pathways may also be related to disease processes. The disease-related pathway may be an intrinsic process of cells in humans in which a pathology arises from, for instance mutation, deletion or under or over expression of a gene. Alternatively the biochemical pathway may be one that is specific to a pathogenic organism or the mechanism of host invasion. In this case, component proteins of such processes may be targets of a therapeutic strategy, such as development of drugs that inhibit invasion by the organism or a component enzyme in a biochemical pathway specific to the pathogenic organism.

Inflammatory diseases are a case in point that can concern both examples. The protein-protein interactions that mediate the adhesion of leukocytes to inflamed tissues are known to involve such proteins as vascular cell adhesion molecule-1 (VCAM-1), and certain cytokines such as IL-6 and IL-8 that are produced during inflammation. However, many of the proteins involved in onset of inflammatory response remain unknown; further, the intracellular signaling pathways triggered by the extracellular associations are poorly understood. The PCAs could be used in elucidation of the mechanisms underlying the onset of inflammation, as well the ensuing signaling. For example, signaling pathways associated with inflammation, such as those mediated by IL-1, IL-6, IL-8 and tumor necrosis have been studied in some

detail and many direct and downstream regulators are known. These regulators can be used as starting point targets in a PCA screening to identify other signalling or modulating proteins that could also be targets for drug development.

There is an increased risk of infection by enteric pathogens in the occurrence of the intestinal inflammation that characterizes idiopathic intestinal diseases. There are two mechanisms which need to be better understood here and which can be addressed by PCA:

i) the cellular mechanisms of inflammation as described above, and ii) the discovery of the specific cell-surface ligands which the pathogenic organisms recognize and associate with. Secreted proteins produced by the pathogen can bind to the basolateral membrane of epithelial cells (as in the case in *Yersinia pseudotuberculosis* infection) or be translocated into intestinal epithelial cells (*Salmonella* infection), promoting infectivity and/or physiological responses to the infection. However, in most cases the interactions between the pathogenic protein and the epithelial cells are unknown.

Cell adhesion and nervous system regeneration

A related example in cell adhesion includes processes involved in develoment and regeneration in the nervous system. Cadherens are membrane proteins that mediate calcium dependent cell-cell adhesion. To do so they need another class of cytoplasmic proteins called cathenins. Those make a bridge between cadherins and cytoskeleton. Cathenins also regulate genes that control differentiation-specific genes. For instance, the protein B-cathenin can interact in certain situation with a transcription factor (lef-1) and be translocated into the nucleus where it constrains the number of genes transactivated by lef-1 (differentiation). This process is regulated by the Wnt signaling pathway (homologs to the wingless pathway in drosophila) by inactivation of GSK3B which permit degradation of APC (a cytoplasmic adapter protein). PCA strategies could be used to identify novel proteins involved in the regulation of these processes..

Proteins involved in viral integration processes are examples of targets that could be tested for inhibitors using the PCA strategies. Examples for the HIV virus include:

i) inhibition of integrase or the transport of the pre-integration complex: protein Ma or vpr.

ii) Inhibition of the cell cycle in G2 by vpr (interaction by cyclin B) causing induction of apoptosis.

iii) Inhibition of the interaction of gp160 (precursor of the membrane proteins) with furine.

Accessory proteins of HIV as a therapeutic target:

i) Vpr: nuclear localizing sequence (target): interaction site of vpr with phosphatasesA .

ii) vif: interaction with vimentin (cytoskeleton associated protein).

ii) Vpu: Degradation of CD4 in the RE mediated by the cytoplasmic tail of Vpu.

iii) nef: Myristoylation signal of Nef.

EXAMPLE 3

Other Examples of Protein Fragment Complementation Assays

Other examples of assays are herein examplified. The reason to produce these assays is to provide alternative PCA strategies that would be appropriate for specific protein association problems such as studying equilibrium or kinetic aspects of assembly. Also, it is possible that in certain contexts (for example, specific cell types) or for certain applications, a specific PCA will not work but an alternative one will. Brief descriptions of each other PCAs embodiments are provided hereinbelow.

1) Glutathione-S-Transferase (GST)

GST from the flat worm *Schistosoma japonicum* is a small (28 kD), monomeric, soluble protein that can be expressed in both prokaryotic and eukaryotic cells. A high resolution crystal structure has been solved and serves as a starting point for design of a PCA. A simple and inexpensive colorimetric assay for GST activity has been developed consisting of the reductiv conjugation of reduced glutathione with 1-chloro-2,4-dinitrobenzine (CNDB), a

brilliant yellow product. A PCA based on similar structural criteria us d to develop the DHFR PCA using GCN4 leucine zippers as oligomerization domains was thus designed. Cotransformants of zipper-GST-fragment fusions are expressed in *E. coli* on agar plates and colonies are transferred to nitrocellulose paper. Detection of fragment complementation is detected in an assay where a glutathione-CDNB reaction mixture is applied as an aerosol on the nitrocellulose and colonies expressing co-expressed fragments of GST are detected as yellow images.

2) Green Fluorescent Protein (GFP)

GFP from *Aequorea victoria* is becoming one of the most popular protein markers for gene expression. This is because the small, monomeric 238 amino-acids protein is intrinsically fluorescent due to the presence of an internal chromophore that results from the autocatalytic cyclization of the polypeptide backbone between residues Ser65 and Gly67 and oxidation of the - bond of Tyr66. The GFP chromophore absorbs light optimally at 395 nm and also possesses a second absorption maximum at 470 nm. This bi-specific absorption suggests the existence of two low energy conformers of the chromophore whose relative population depends on local environment of the chromophore. A mutant Ser65Thr that eliminates isomerization (single absorption maximum at 488 nm) results in a 4 to 6 times more intense fluorescence than the wild type. Recently the structure of GFP has been solved by two groups, making it a candidate for a structure-based PCA-design, which has begun to be developped. As with the GST assay, all of the initial development is done in *E. coli* with GCN4 leucine zipper-forming sequences as oligomerization domains. Direct detection of fluorescence by visual observation under broad spectrum UV light will be used. This system will also be tested in COS cells, selecting for co-transfectants using fluorescence activated cell sorting (FACS).

3) Fire Fly Luciferase

Firefly luciferase is a 62 kDa protein which catalyzes oxidation of the heterocycle luciferin. The product possesses one of the highest quantum yields for bioluminescent reactions: one photon is emitted for every

oxidized luciferin molecule. The structure of luciferase has recently been solved, allowing for strucutre-based development of a PCA. As with the GST assay described herein, cells are grown on a nitrocellulose matrix. The addition of the luciferin at the surface of the nitrocellulose permits it to diffuse across the cytoplasmic membrane and trigger the photoluminescent reaction. The detection is done immediately on a photographic film. Luciferase is an ideal candidate for a PCA: the detection assays are rapid, inexpensive, very sensitive, and utilize a non-radioactive substrate that is available commercially. The substrate of luciferase, luciferin, can diffuse across the cytoplasmic membrane (under acidic pH), allowing the detection of luciferase in intact cells. This enzyme is currently utilized as a reporter gene in a variety of expression systems. The expression of this protein has been well characterized in bacterial, mammalian, and in plant cells, suggesting that it would provide a versatile PCA.

4) Xanthine-guanine phosphoribosyl transferase (XGPRT)

The *E. coli* enzyme XGPRT converts xanthine to xanthine monophosphate (XMP), a precursor of GMP. Because the mammalian enzyme hypoxanthine-guanine phosphoribosyl transferase HGPRT can only use hypoxanthine and guanine as substrates, the bacterial XGPRT can be used as a dominant selection assay for a PCA for cells grown in the presence of xanthine. Vectors expressing XGPRT confer the ability of mammalian cells to grow in selective medium containing adenine, xanthine, and mycophenolic acid. The function of mycophenolic acid is to inhibit *de novo* synthesis of GMP by blocking the conversion of IMP into XMP (Chapman A. B. 1983, Molec. Cell. Biol. 3:1421-1429). The only GMP produced comes from the conversion of xanthine into XMP, catalyzed by the bacterial XGPRT. As with aminoglycoside phosphotransferase fragments of XGPRT can be generated based on the known structure (see table 1) using the design-evolution strategy described above with fragments fused to the GCN4 leucine zippers as a test oligomerization domains. The complementary fusions are cotransfected and the proteins transiently expressed in COS-7 cells, or stability expressed in CHO cells, grown in the selective medium. In the case of CHO cells, colonies are collected and sequentially re-cultured at increasing concentrations of the selective compounds

in order to enrich for populations of cells that efficiently express the fusions at high concentrations.

5) Adenosine deaminase

Adenosine deaminase (ADA) is present in minute quantities in virtually all mammalian cell. Although it is not an essential enzyme for cell growth, ADA can be used in a dominant selection assay. It is possible to establish growth conditions in which the cells require ADA to survive. ADA catalyzes the irreversible conversion of cytotoxic adenine nucleosides to their respective nontoxic inosine analogues. By adding cytotoxic concentrations of adenosine or cytotoxic adenosine analogues such as 9-b-D-xylofuranosyladenine to the cells, ADA is required for cell growth to detoxify the cytotoxic agent. Cells that incorporate the ADA gene can then be selected for amplification in the presence of low concentrations of 2'-deoxycoformycin, a tight-binding transition state analogue inhibitor of ADA. ADA can then be used for a PCA based on cell survival (Kaufman et al. 1986, Proc. Nat. Acad. Sci. (USA) 83:3136-3140). As with the other systems described above, fragments of ADA can be generated based on the known structure (see table 1) using the design-evolution strategy described above with fragments fused to the GCN4 leucine zippers as a test oligomerization domains. The complementary fusions are cotransfected and the proteins transiently expressed in COS-7 cells, or stability expressed in CHO cells, grown in the selective medium containing 2'-deoxycoformycin. In the case of CHO cells, colonies are collected and sequentially re-cultured at increasing concentrations of 2'-deoxycoformycin in order to enrich for populations of cells that efficiently express the fusions at high concentrations.

6) Bleomycin binding protein (zeocin resistance gene)

Zeocin, a member of the bleomycin/phleomycin family of antibiotics, is toxic to bacteria, fungi, plants, and mammalian cells. The expression of the zeocin resistance gene confers resistance to bleomycin/zeocin. The protein confers resistance by binding to and sequestering the drug and thus preventing its association and hydrolysis of DNA (Berdy, J. 1980, In Amino Acid and Peptide Antibiotics, Berdy, ed. Boca Raton,

FL: CRC Press, pp.459-497; Mulsant et al. 1989, Somat. Cell. Mol. Genet. 14:243-252). Bleomycin binding protein (BBP) could then be used for a PCA based on cell survival. As with the other systems described above, fragments of ADA can be generated based on the known structure (see table 1) using the design-evolution strategy described above with fragments fused to the GCN4 leucine zippers as a test oligomerization domains. The BBP is a small (8 kD) dimer that binds to drugs *via* a subunit interface binding site. For this reason, the design would be somewhat different in that first, a single chain form of the dimer would be generated by making a fusion of two BBP genes with a short sequence coding for a simple polypeptide linker introduced between the two subunits. Fragments in this case will be based on a short sequence of one of the subunit modules, while the other fragment will be composed of the remaining sequence of the subunit plus the other subunit. Complementation and selection experiments will be performed as described for the examples above using bleomycin or zeocin as selective drugs.

7) Hygromycin-B-phosphotransferase

The antibiotic hygromycin-B is an aminocyclitol that inhibits protein synthesis by disrupting translocation and promoting misreading. The *E. coli* enzyme hygromycin-B-phosphotransferase detoxifies the cells by phosphorylating Hygromycin-B. When expressed in mammalian cells, hygromycin-B-phosphotransferase can confer resistance to hygromycin-B (Gritz et al. 1983, Gene 25:179-188). The enzyme is a dominant selectable marker that could be used for a PCA based on cell survival. While the structure of the enzyme is not known it is suspected that this enzyme is homologous to aminoglycoside kinase (Shaw et al. 1993, Microbiol. Rev. 57:138-163). It is therefore possible to use the combined design/evolution strategy disclosed herein to produce a PCA with this enzyme and perform dominant selection in mammalian cells with selection at increasing concentrations of hygromycin B.

8) L-histidinol NAD+oxydoreductase

The hisD gene of *Salmonella typhimurium* codes for the L-histidinol NAD+oxydoreductase that converts histidinol to histidine. Mammalian cells grown in media lacking histidine but containing histidinol can be selected

for expression of hisD (Hartman et al. 1988, Proc. Nat. Acad. Sci. (USA) 85:8047-8051). An additional advantage of using hisD in dominant selection is that histidinol is itself toxic, inhibiting the activity of endogenous histidyl-tRNA synthetase. Furthermore, histidinol is inexpensive and readily permeates cells. The structure of histidinol NAD+oxydoreductase is unknown and so development of a PCA based on this enzyme is based entirely on the exonuclease fragment/evolution strategy, disclosed above.

The following Table list alternative embodiments using other PCA reporters. Abreviations in Table: Type: D, dominant selection marker; R, recessive selection marker. Structure: four letter codes= Protein Data Bank (PDB) entries; K, known but not deposited in PDB; U, unknown. mono/oligo: M, monomer; D, dimer; tetra, tetramer.

TABLE 1. A list of Oth r P tential PCA R porter Candidates

A-Assays based on Dominant or Recessive Selection

Enzyme	Type	Structur e	Size	mono/ oligo	Selection drugs/Conditions
DHFR	R/D	many	18kD	M	methotrexate/trimethoprim
Adenosine deaminase	D/R	1ADD		M	Xyl-A or adenosine,alanosine, and 2'-deoxycoformycin
Thymidine kinase	D/R	1KIN		D	gangcyclovene, HAT
Mutant hypoxanthine- guanine phosphoribosyl transferase	D	1HGM		D	HAT + thymidine kinase
Thymidylate synthetase	R	1NJE	35kd	M	2 fluorodeoxyuridine
Xanthine-guanine phosphoribosyl transferase	D	1NUL			mycophenolic acid with limiting xanthine
Glutamine synthetase	R	2LGS			
Asparagine synthetase	R	U			B-aspartyl hydroxamate or albizin
Puromycin *N*-acetyltransferase	D	U	23kD	M	puromycin
Aminoglycoside phosphotransferase	D	K	35kD	M	neomycin, G418, gentamycin
Hygromycin B phosphotransferase	D	U		M	hygromycin B
L-histidinol:NAD+ oxidoreductase	D	U	46kD	M	histidinol
Bleomycin binding protein	D	K	8kD	D	bleomycin/zeocin
Cytosine methyl-transferase	R/D	U			5-Azacytidine (5-aza-CR) and 5-aza-2'-deoxycytidine
O6-alkylguanine alkyltransferase	D	1ADN			*N*-methyl-*N*-nitrosourea
Glycinamide ribonucleotide transformylase	R	1GRC	23.2kD	D	dideazatetrahydrofolate, minus purine
Glycinamide ribonucleotide synthetase	R	U	45.9kD		minus purine
Phosphoribosyl-aminoimidazole synthetase	R	U	36.7kD		minus purine
Formylglycinamide ribotide amidotransferase	R	U	141.4kD	M	L-azaserine, 6-diazo-5-oxo-L-norleucine, minus purine
Phosphoribosyl-aminoimidazole carboxylase	R	U	39.5kD	D	minus purine
Phosphoribosyl-aminoimidazole carboxamide formyltransferase	R	U	57.3kD		minus purine
Fatty acid synthase	R		272kD	D	cerulenin
IMP dehydrogenase	R	1AK5	55.4kD	Tetra	mycophenolic acid

ii-Viral Plaque Assays

Enzyme	Type	Structure	Size	Mono/ Oligo	Selection drugs/Conditions

Thioredoxin	D	1TDF	34.5 kD	D	
Reverse transcriptase	D	3HVT			
Viral protease	D				

B-Cell Death Assays

Enzyme	Type	Structure	Size	Mono/ Oligo	Selection drugs/Conditions
Cysteine protease: papain	D	1STF	38.9 kD	M	inhibited by cystatin
Cysteine protease: caspase	D	1CP3	17kD + 12kD	HeteroD	inhibited by DEVD-aldehyde (can also by used in a fluorimetric or colorimetric assay, *in vitro*)
Metalloprotease: carboxypeptidase	D		47.1 kD	M	inhibited by methyl-ethyl succinic acid
Serine protease: proteinase K	D	1PTK	30.6 kD	M	inhibited by serpins
Aspartic protease: pepsin	D	1PSN	34.5 kD	M	inhibited by pepstatin A (can also be used in an fluorimetric assay, *in vitro*)
Lysozyme	D	many	23.2 kD	M	inhibited by N-acetylglucosamine trisaccharide
RNAse	D	many	13.3 kD	M	inhibited by RNAse inhibitor
DNAse	D	1DNK	61.6 kD	M	inhibited by actin
Phospholipase A2	D	1P2P	13.8 kD	M/D	many inhibitors: bromophenacyl bromide, hexadecyl-trifluoroethyl-glycero-phosphomethanol, bromoenol lactone, etc.
Phospholipase C	D	1AH7	28kD	M	many inhibitors: neomycin, chelerythrine, U73122, etc.

C-Colorimetric/Fluorimetric Assay

Enzyme	Structure	Size	Mono/ Oligo	Selection drugs/Conditions
DT-Diaphorase (NAD(P)H-[quinone acceptor] oxidoreductase)	1QRD	26kD	D	NADPH-diaphorase stain, inhibited by dicumarol, Cibacron blue and phenidione Note: can also be used in a cell death assay (+nitrobenzimidazole, fo example).
(NAD(P)H-[quinone acceptor] oxidoreductase)-2	isoform of 1QRD	21kD	D	NRH-diaphorase stain, inhibited by pentahydroxyflavone
Thermophilic diaphorase (*Bacillus stearothermophilus*)		30kD	M	NADH-diaphorase stain
Glutathione-S-transferase	1GNE	26kD other isoform of 28kD	D	production of a yellow product by the conjugation of glutathione with an aromatic substance, chloro dinitrobenzene (CDNB)
Luciferase	1LCI	62kD	M	Fluorometric
Green-fluorescent protein	1EMA	30kD	M	Intrinsic fluorescence
Chloramphenicol acetyltransferase	1CLA	25kD	Tri	Fluorimetric: Bodipy chloramphenicol
Uricase		32kD	Tetra	Fluorometric
SEAP (secreted form of human placental alkaline phosphatase)	1AJA		M	CSPD chemiluminescent substrate
B-Glucuronidase	1BHG	71kD	Tetra	Histochemical, fluorometric or spectrophotometric assays using various substrates such as X-GLUC.

D-Heteromeric Enzyme Strategies

Tyrosinase		30kD + 14kD	Hetero M+M	Colorimetric: synthesis of melanin

EXAMPLE 4

Examples of Variants of PCA to detect multiple protein/protein-dna/protein RNA/protein-drug complexes

While specific examples have only been given for applications of PCA to protein-pair interactions, it is possible to apply PCA to multiprotein, protein-RNA, protein-DNA or protein-small molecule interactions. There are two general schemes for achieving such systems. Multi-subunit PCA: Two proteins need not interact for a PCA signal to be observed; if a partner protein or protein complex binds to two proteins simultaneously, it is possible to detect such a three protein complex. A multusubunit PCA is conceived with the example of herpes simplex virus thymidine kinase (TK), a homodimer of 40 kD . In this conception, the TK structure contains two well defined domains consisting of an alpha/beta (residues 1-223) and an alpha-helical domain (224-374). As a test system, the Rop1 dimer, a four helix bundle homodimer was used. The two fragments of TK are extracted by PCR and subcloned into the transient transfection vector pMT3, the first in tandem to the Adenovirus major late promoter, tripartite leader 3' to the first ATG, and the second downstream of a ECMV internal initiation site. Restriction sites previously introduced between the first and the last ATG are subcloned into BamHI/ KpnI and PstI/EcoRI cloning sites downstream of the two ATGs. These are used to subclone PCR-generated fragments of the Rop1 subunits into two different vectors. Subsequently Ltk- cells are cotransfected by lipofection with the two plasmids and colonies of surviving cells are serially selected in medium containing increasing concentrations of HAT (hypoxanthine/ aminopterin/thymidine). Cells that express complementary fragments of TK fused to the four Rop1 will proliferate under this selective pressure, or otherwise die. Specific examples of use of this concept would be in determining constituents of multiprotein complexes that are formed transiently or constitutively in cells.

The utility of PCA is not limited to detecting protein-protein interactions, but can be adapted to detecting interactions of proteins with DNA, RNA, or small molecules. In this conception, two proteins are fused to PCA complementary fragments, but the two proteins do not interact with each other. The interaction must be triggered by a third entity, which can be any molecule that will simultaneously bind to the two proteins or induce an interaction between the two proteins by causing a conformational change in one or both of the

partners. Two examples have been demonstrated by the Applicant using the mDHFR PCA in *E. coli*. In the first case a natural product, the immunosuppressant drug rapamycin, is used to induce an interaction between its receptor FKBP12 and a partner protein mTOR (mammalian Target of Rapamycin). The interaction was detected by cotransformation of DHFR fragments fused to FKBP or mTOR into *E. coli* grown in the presence or absence of trimethoprim (as described above) and rapamycin (0- 10 nM). Support of growth as detected by colony formation was demonstrated to be completely dependent on the addition of rapamycin, suggesting that the mDHFR PCA is detecting a rapamycin-induced assembly of a FKBP12-mTOR and subsequent reconstitution of DHFR activity. This is one example of a use of the PCA strategy to test for small molecules that can induce interactions between proteins. General applications could be made for therapeutic development. For example, screening small molecule combinatorial compound libraries for molecules that induce interactions between proteins could be carried out. These molecules may inhibit the activities of either or both of the proteins, or activate specific cellular processes that are initiated by other events, such as growth factor-mediated receptor dimerization. The discovery of such small molecules could lead to the development of orally available drugs for the treatment of a broad spectrum of human diseases.

Another example of an induced interaction that was studied with the DHFR PCA is the interaction of the oncogene GTPase p21 ras and its direct downstream target, the serine/threonine kinase raf. This interaction only occurs when the GTPase is in the GTP-bound form, whereas turnover of GTP to GDP leads to release of the complex. As with the FKBP-mTOR complex, this induced interaction in *E. coli* could be demonstrated. PCA could be used in a general way to study such induced interactions, and to screen for compounds that release or prevent these interactions in pathological states. The ras-raf interaction itself could be a target of therapeutic intervention. Oncogenic forms of ras consist of mutants that are incapable of turning over GTP and therefore remain continuously associated with activated ras. This leads to a constitutive uncontrolled growth signal that results, in part, in oncogenesis. The

identification of compounds that inhibit this process, by PCA, would be of value in broad treatment of cancers. Other examples of multimolecular applications of PCA could include identification of novel DNA or RNA binding proteins. In its simplest conception, a skilled artisan could use a known DNA or RNA binding motifs, for instance a retinoic acid receptor zinc finger, or a simple RNA binding protein such as IF-1, respectively. One half of the PCA could consist of the DNA or RNA protein binding domain fused to one of the PCA fragments (control fragment). The complementary fragment would be fused to a cDNA library. A third entity, the gene encoding a sequence containing an element known to bind to the control protein, and then a second putative or known regulatory element is coded for after this sequence could be used. A test system consists of tat/tar elements that control elongation in transcription/translation of HIV genes. An example of an application would be the identification of tat binding elements that have been proposed to exist in eukaryotic genomes and may regulate genes in the same or similar way to that of HIV genes. (SenGupta et al. 1996, Proc. Natl. Acad. USA 6:8496-8501).

EXAMPLE 5

Examples of PCA applications to drug screening: Screening combinatorial libraries for compounds that inhibit or induce protein-protein/protein-RNA/protein-DRNA complexes

A) Drug screening.

The PCA strategy can be directly applied to identifying potential therapeutic molecules contained in combinatorial libraries of organic molecules. It is possible to perform high throughput screening of such libraries to screen for compounds that will inhibit or induce protein-protein interactions or protein-nucleic acid interactions (as discussed above). In addition it is also possible to screen for compounds that inhibit enzymes whose substrates are other proteins, DNA, RNA or carbohydrates, as discussed below. In this application, proteins that interact/protein substrate pairs, or control DNA/RNA binding protein-enzyme pairs are fused to PCA complementary fragments and plasmids harboring these pairs are transformed/transfected into a cell, along with any third DNA or RNA element as the case requires. Transformed/transfected cells are grown in liquid cultures in multiwell plates, each well being inoculated with a single compound from an array of combinatorially synthesized compounds. A readout of a response depends on the effect of a compound. If the compound inhibits a protein interaction, there is a negative response (no PCA signal is the positive response). If the compound induces a protein interaction, the response is a positive PCA signal. Controls for non-specific effects of compounds include: 1) demonstration that the compound does not effect the PCA enzyme itself (test against cells transfected with the wild-type intact enzyme used as the PCA probe) and in the case of a cell survival assay, that the compound is not toxic to the cells that have not been transformed/transfected. As well as providing a high throughput assay for biological activity of compounds, PCA also offers the advantage over *in vitro* assays that it is a test for cell membrane permeability of active compounds. Specifically demonstrated examples of PCA for drug screening from the Applicant include the application of DHFR PCA in *E. coli* for the detection compounds that inhibit therapeutically relevant targets, such as

Bax/Bcl2, fkbp12/tor, ras/raf, the carboxy-terminal dimerization domain of HIV-1 capsid protein, IkB kinase IKK-1 and IKK-2 dimerization domains (leucine zippers and helix-loop-helix domains). In each case, the two proteins are subcloned 5' upstream of either F[1,2] or F[3] as described above. Plasmids harboring the complementary fragments are cotransformed into BL21 cells. Colonies from minimal medium plates containing IPTG and trimethoprim are picked, and grown in liquid medium under the same selective conditions and frozen stocks made. For a single screening cycle, a priming overnight culture is grown from frozen stocks in LB medium. A selective minimal medium containing trimethoprim, ampicillin, IPTG is aliquated at 25 ml into each well of a 384 well plate. Each well is then inoculated with 1 ul of an individual sample from a compound array (ArQule Inc.) to give a final concentration of 10 uM. Each well is then inoculated with 2 ml of overnight culture and plates are incubated in a specially adapted shaker bath at 37°C. At 2 hour intervals, plates are read on an optical absorption spectroscopic plate reader coupled to a PC and spreadsheet software at 600 nm (scattering) for a period of 8 hours. Rates of growth are calculated from individual time readings for each well and compared to a standard curve. A "hit" is defined as the identification of an individual compound which reduces the rate of growth to less than the 95 % confidence interval based on the standard deviation for growth rates observed in all of the wells within the test plate. "Near hits" are defined as those cases where growth rates are within the 95 % confidence interval. For each of the hits or near hits, the following controls are then performed: The same experiment is performed with BL21 cells that are transformed with empty vector (and no trimethoprim), with vector harboring the full length mDHFR gene, or with cotransfected cells where protein expression is not induced by IPTG. If in all of these cases the compound has no effect, it can be concluded that it is specifically disrupting the protein-protein interaction being tested. Such validated hits or near hits are then retested to establish a dose-response curve for the individual compound, with concentrations varying from 1 pM up to 1 μM by orders of magnitude of 10. The PCA strategy for compound screening can also be applied in the multiprotein protein-RNA/DNA cases as described above,

and can easily be adapted to the DHFR or any other PCA in *E. coli* or in yeast versions of the same PCAs. Such screening can also be applied to enzymes whose targets are other proteins or nucleic acids for known enzyme/substrate pairs or to novel enzyme substrate pairs identified as described below.

Proteins involved in viral integration processes are examples of targets that could be tested for inhibitors using the PCA strategies. Examples for the HIV virus and accessory proteins of HIV as a therapeutic target have been given in Example 2.

Other general targets for drug screening could include proteins linked to neurodegenerative diseases, such as alpha-synuclein. This protein has been linked to early onset of Parkinson's disease and it is present also implicated in in Alzheimer's disease. Another example is β-amyloid proteins, also linked to Alzheimer's disease.

An example of protein-carbohydrate interactions that could be a target for drug screening includes the selectins that are generally implicated in inflammation. These cell surface glycoproteins are directly involved in diapedesis.

A number of tumor supressor genes whose actions are mediated by protein-protein interactions could be screened for potential anti-cancer compounds. These include PTEN, a tumor supressor directly involved in the formation of harmatomas, in inherited breast and in thyroid cancer. Other interesting tumor supressor genes include p53, Rb and BARC1.

EXAMPLE 6

Examples of applications of the PCA strategy to detect enzyme/substrate interactions

The examples described above are used for identifying novel molecular interactions involving molecules that merely bind to each other. However detecting the substrates of enzymes is also fully compatible with the PCA strategy as shown below:

i) Enzymes that form tight complexes with protein substrates or induce efficient PCA fragment assembly or

ii) Mutant enzymes that bind tightly to substrates but do not undergo product release because of mutations residues involved in nucleophilic attack and/or product release (substrate trapping).

Enzymes may form tight complexes with their substrates (Kd ~1-10 μM). In these cases PCA may be efficient enough to detect such interactions. However, even if this is not true, PCA may work to detect weaker interactions. Generally, if the rate of catalysis and product release is slower than the rate of folding- reassembly of the PCA complementary fragments, effectively irreversible folding and reconstitution of the PCA reporter activity will have occurred. Therefore, even if the enzyme and substrate are no longer interacting, the PCA signal can be detected. Therefore, the detection of novel enzyme substrates using PCA may be possible, independent of effective substrate Kd or rate of product release. In cases where product release is much faster than PCA fragment assembly/folding, an alternative approach is provided by generating "substrate trapping" mutants of the test enzyme. An example of this approach applies to the protein tyrosine phosphatase PTP1B, for which substrate trapping mutants have been generated by mutating the nucleophilic aspartate 181 to alanine, rendering the enzyme catalytically dead, but capable of forming tight complexes with a known substrate, the EGF receptor and other unknown proteins (Flint et al. 1996, Proc. Natl. Acad. USA 94:680-1685).

An application of using PCA to screen for interacting partners of PTP1B is given as follows. First, the aminoglycoside kinase (AK)-based PCA in transiently transfected COS or 293 cells is used. The substrate trap mutant catalytic domain of PTP1B is fused to *N*-terminal complementary fragment of AK, while a C-terminal fusion of the other AK fragment is made to a cDNA library. Cells are co-transfected with complementary AK pairs and grown in selective concentrations of G418. After 72 hours, colonies of surviving cells are picked and *in situ* PCR is performed using primers designed to anneal to 3' and 5' flanking regions of the cDNA coding region. PCR amplified products are then 5' sequenced to identify the gene.

Enzyme inhibitors Screening combinatorial libraries of compounds for those that inhibit enzyme-PROTEIN substrate complexes can also be carried out with:

i) Enzymes that form tight complexes with protein substrates; or

ii) Mutant enzymes that bind tightly to substrate but do not undergo product release because of the mutation.

EXAMPLE 7

Applications of the PCA strategy to protein engineering/evolution The PCA strategy can be used to generate peptides or proteins with novel binding properties that may have therapeutic value, as is done with phage display technology. It is also possible to develop enzymes with novel substrates or physical properties for industrial enzyme development. Two detailed examples of such applications of the PCA strategy are, with additional applications listed below.

1) Selection of high-affinity, heterodimerizing leucine zipper sequences (J. Pelletier, K. Arndt, A. Plueckthun and S. Michnick, manuscript in preparation)

The mDHFR PCA, described above, was used in a scheme for the selection of efficiently heterodimerizing, designed leucine zippers. It has been proposed that the formation of salt bridges between positively and negatively charged residues at complementing "e" and "g" positions is important in stabilizing leucine zipper formation, though this view has been contested. In order to help define the importance of salt-bridge formation at the e and g positions, two leucine zipper libraries were built. Both are based on the GCN4 leucine zipper sequence, but contain sequence information specific to either Jun or Fos zippers in order to create heterodimerizing pairs. As well, the e-1 to e-4 and g-1 to g-4 positions in each library were randomized to code for positively or negatively charged residues, or neutral polar residues. These libraries were amplified by PCR and subcloned into the Z-F[1,2] or Z-F[3] constructs (described above) from which the GCN4 zipper sequences had been removed. The bacterial mDHFR PCA selection was performed on selective solid media, as described earlier. Colonies were picked and sequenced; sequence analysis

reveals that the distribution of charged or neutral residues at e-g pairs is not random, but is bias d toward pairing of opposite charges, or pairing of a charged with a neutral residue, rather than same-charge pairing (see figure 7). It was reasoned that better zipper pairing should lead to an increase in efficiency of DHFR-fragment complementation, resulting in faster bacterial doubling times (see Table 1 in the mDHFR PCA description), was thus undertaken and undertook a selection/enrichment of the novel zippers relative to GCN4, as follows. The designed zipper libraries, expressed as *N*-terminal fusions to the DHFR F[1,2] or F[3:I114A], were cotransformed, clones were picked, propagated and mixed in selective liquid culture, and the mix was added in a 1:1 000 000 ratio to clone Z-F[1,2] + Z-F[3:I114A] (original GCN4 leucine zippers). The mixture was propagated in selective liquid culture over multiple passages. Restriction analysis shows that within 4 passages, the population of GCN4-expressing bacteria is diminishing relative to the novel zipper sequences (data not shown), indicating that some of the designed zipper-containing clones are propagated at a higher rate than those containing GCN4. Bacteria from later passages were plated on selective medium, and individual clones sequenced to reveal the identity of the most successful by designed zipper pairs (data not shown).

2) Application of PCA to enzyme function and design

PCA Development: Adenosine deaminase (ADA) meets all of the criteria for a PCA listed above. ADA is a small (~40 kD), and easily purified monomeric zinc metallo-enzyme and the structure of murine ADA has been resolved. Several *in vitro* ADA activity assays have been developed, involving UV spectrophotometry and stopped-flow fluorimetry. *E. coli* ADA catalyzes the irreversible conversion of cytotoxic adenine nucleosides to non-toxic inosines. Eukaryotic or prokaryotic cells propagated in the presence of cytotoxic concentrations of adenosine or adenosine analogs require ADA to detoxify these compounds. This is the basis of a dominant-selection strategy used to select for cells expressing a specific gene in mammalian cells. The ADA gene has also been expressed in SF3834 *E. coli* cells which lack a gene coding for endogenous ADA. When the gene coding for ADA is introduced into ADA-

bacterial DNA, those cells that express ADA are able to survive high concentrations of added adenosine; those that do not, die . This forms the basis of an *in vivo* ADA activity assay.

ADA was thus chosen, principally because it can be used as a dominant selective marker in mammalian and bacterial cells in which the gene has been knocked out. The reason a dominant selective gene was chosen is because in screening for novel protein-protein interactions, particularly testing for interactions of a known protein against a library of millions of independent clones, selection serves to filter for cells that may show a positive response for reasons independent with a specific protein-protein interaction. Three test systems of interacting proteins including leucine zipper-forming sequences will thus be used, the proteins raf and p21 and the induced oligomerization system, FK506 binding protein (FKBP) and mTOR that interact through the macrocyclic immuno-suppressant compound rapamycin. For all of these systems, an *E. coli*. construct and mammalian transient transfection plasmids will be used. The test proteins will be subcloned as fusions to ADA complementary fragments. The primary assay will be survival of SF3834 *E. coli* cells that have been transformed with the complementary ADA fragments fused to the test oligomerization proteins in the presense of toxic concentrations of adensosine. Fusion proteins will then be purified from colonies and *in vitro* assays of ADA activity performed as described below. The utility of the ADA PCA as a method to identify novel proteins that interact with a test bait will be performed in mammalian COS-7 and HEK-293T cells transiently transfected with FKBP fused to one of the ADA fragments and the other fragment fused to a cDNA library from normal human spleen containing 10^6 independent clones. As with the *E. coli* assay, cells that survive in a medium containing toxic concentrations of ADA are collected and isolated plasmids will be tested to identify the gene for the interacting protein by PCR amplification and chain propagation-termination techniques.

Structural motifs required for protein function

Determination of the structural elements required for the enzymatic function of ADA are investigated through alteration of the structures of the enzyme fragments. At first, ADA is cut into two separate domains - one

responsible for substrate binding (residues 1-210) and one responsible for catalysis (residues 211-352). These separate pieces will be attached to known assembly domains, such as leucine zippers (see Example 1 above). Reassembly will restore activity which will be assessed through detailed *in vitro* kinetic analysis of the binding and catalytic properties of the re-assembled enzyme, using UV spectrophotometry and stopped-flow fluorimetry to observe the enzymatic reactions. This system will provide another handle on the manipulation of enzyme activity that will afford a powerful tool for enzymatic mechanism study. For example, the difference in the kinetic behaviour of the reassembled enzyme on mixing with the substrate, compared to enzyme reassembled in presence of substrate (where substrate may already be bound by binding domain) will allow a sophisticated level of study of the importance of binding energy to catalysis. Subsequent point mutations to the functional or assembly domains of the proteins will then allow a very subtle perturbation and detailed quantification of the relationship of binding energy to catalysis. This precise control over the structure and assembly of separate functional domains of the enzyme will permit very sophisticated enzymatic structure function studies, the definition of structural motifs and an understanding of their role in catalysis.

Novel protein catalyst design

The detailed knowledge of the enzyme mechanism gained through determination of the structural requirements for catalysis will then be exploited through the combination of these functional "building blocks" with the functional motifs responsible for substrate binding and catalysis in other enzymes, allowing the generation of novel protein catalysts. For example, the catalytic motif from ADA is modified to a cytidine-binding motif, creating a novel enzyme with potentially useful catalytic properties. The activity of these novel enzymes can easily be assessed through *in vivo* assays similar to that of the PCA system, or through *in vitro* activity assays. Furthermore, the detailed mechanistic investigation of the resulting enzymes, possible with this system, will permit the rational design of each subsequent generation of catalysts.

EXAMPLE 8

Examples of applications of the PCA strategy to detect mol cular interactions in whole organisms

The use of the PCA to detect molecular interactions in whole organisms is a logical extension of the PCA applications described above. Whole model organisms such as drosophila, nematodes, zebra fish and puffer fish are non-limiting examples of such organisms. The sole differences with other listed examples is that vectors used would need to be different (for example retroviral vectors) and that any substrates needed by the PCA would need to be bioavailable, or detection would need to be performed *in situ.*

EXAMPLE 9

Examples of applications of the PCA strategy to Gene Therapy

Another important embodiment of the invention is to provide a means and method for gene therapy of mammalian disease. Of particular interest is the use of PCA therapeutic for treatment of cancer. In one embodiment of the PCA gene therapy, a PCA is developed employing fragments (modular protein units) derived from a protein toxin for example: Pseudomonas exotoxin, Diptheria toxin and the plant toxin gelonin, or other like molecules. For therapy of breast cancer for example, a mammalian, retroviral, adenoviral, or eukaryotic artificial chromosomal (EAC's) genetic construct is first prepared. The construct introduces one fragment of the selected toxin under the control of the promoter for expression of the *erb*B2 oncogene. It is well known that the *erb*B2 oncogene is overexpressed in breast cancer and adenocarcinoma cells (Slamon et. al., Science, 1989, 244:707). The HER2/*neu* (*c-erb*B-2) proto-oncogene encodes a sub-class 1 185-kDa transmembrane protein tyrosine kinase growth factor receptor, $p185^{HER2}$. Also, the human *erb*B2 oncogene is located on chromosome 17, region q21 and comprises 4,480 base pairs and $p185^{HER2}$ serves as a receptor for a 30-kDa glycoprotein growth factor secreted by human breast cancer cell lines (Lupu et al., Science, 1990, 249:1152).

The transgene is introduced *in vivo* or *ex-vivo* into target cells employing methods known by those skilled in the art (e.g. homologous

recombination to insert transgene into the locus of interest via retroviral, adenoviral or EAC's). A second genetic construct comprises a fusion gene containing a target DNA that encodes an interacting protein that interacts with the *erb*B2 oncogene (discovered by the PCA process described in this invention) and the "second" fragment of the toxin molecule. This construct is delivered to the patient by methods known in the art. For example, the construct can be delivered as shown in U.S. Patent Nos. 5,399,346 and 5,585,237 whose entire contents are incorporated by reference herein. Transgene expression of the *erb*B2 oncogene-toxin fragment described will now be under the control of the constitutive oncogene promoter. Proliferating tumor cells will thus produce one piece of the toxin attached as a fusion to the *erb*B2 oncogene. In the presence of the second genetic construct expressing the PCA discovered interacting *erb*B2 oncogene "interacting protein - toxin fragment" construct then: *erb*B2 oncogene-toxin fragmentA: interacting protein-toxin fragment B will be created and induce death of target tumor cells through the creation of an active toxin through Protein Fragment Complementation and thus provide an efficacious and efficient therapy of the disease.

This can be extended to other diseases and other toxins employing techniques described and embodied in this invention.

EXAMPLE 10

Examples of applications of the PCA strategy to detect molecular interactions *in vitro*

Any of the PCA strategies described above could be adapted to *in vitro* detection. Unlike the *in vivo* PCAs however, detection would be performed with purified PCA fragment-fusion proteins. Such uses of PCA have the potential for use in diagnostic kits. For example the test DHFR assay described above where the interactiing domains are FKBP12 and TOR could be used as a diagnostic test for rapamycin concentrations for use in monitoring dosage in patients treated with this drug.

EXAMPLE 11

Signaling by the Erythropoietin Receptor Mediated by a Ligand-induced Conformation Change in Constitutive Receptor Dimers

The instant Example illustrates a fluorescent assay based on a dimerization-induced complementation of designed fragments of the enzyme murine dihydrofolate reductase (DHFR). The basis for the assay is that complementary fragments of DHFR when expressed and reassembled in cells, will bind to the high affinity (Kd= 100 pM) fluorescein-conjugated inhibitor methotrexate (fMTX) in a 1:1 complex. fMTX is retained in cells by this complex, while unbound is actively and rapidly transported out of the cells . In addition, binding of fMTX to DHFR results in an 4.5 fold increase in quantum yield. Bound fMTX, and by inference reconstituted DHFR, can then be monitored by fluorescence microscopy, FACS or spectroscopy. Since the complex of fMTX with DHFR is 1:1, measured fluorescence can be calibrated to determine average numbers of complexes in individual cells or averages in a population of cells. To test the allosteric model of receptor activation it was reasoned as thate: if the receptor transmembrane domain is separated by the distance observed in the crystal structure of unligated EpoR, then DHFR fragments fused to the C-terminal of the transmembrane domains will complement only if ligand induces the necessary conformation change that allows the fragments to come into contact. Furthermore, the absolute regio- and stereospecific requirement that fragments be sufficiently close to fold-reassemble into the enzyme three dimensional structure means that a false response, merely proximal due to interacting proteins, is unlikely. In addition, insertion of flexible linker peptides of a critical length between the transmembrane domain and the fragments should result in constitutive complementation, insensitive to ligand. Based on the EpoR crystal structure, the minimum length of linker necessary for a constitutive response would be 10 amino acids, assuming that the length of an average peptide bond is ~4 Å and the distance separating the fragments is 82 Å. Longer linkers should result in complementation, independent of ligand. Linkers of 5, 10 and 30 amino acids, corresponding to extended lengths of 20, 40, and 120 Å, respectively, were thus used.

CHO DUKX-B11 (DHFR$^-$) cells were co-transfected with EpoR extracellular and transmembrane domains fused to the variable linkers and one of the two DHFR complementary fragments F[1,2] or F[3]. Co-transfectants were selected in nucleotide-free medium (selection for DHFR activity) and in the presence of Epo (2 nM) to assure that activated receptor and therefore complementation and reconstitution of DHFR activity was present. Fluorescence microscopy (Fig. 9) of unfixed, co-transfected cells, that had been incubated with fMTX showed high levels of fluorescence (no nuclear fluorescence was observed) when cells were pretreated with Epo or with the EpoR agonist peptide EMP1 at saturating concentrations. In the absence of ligands, cells transfected with EpoR-DHFR fragment fusions connected by 5 amino acid linkers showed no fluorescence, compared to non-transfected cells, those with 10 amino acid linkers showed a small background of constitutive fluorescence, but those cells expressing fusions with a 30 amino acid linker showed the same level of fluorescence in the presence or absence of ligands. These results were confirmed by FACS analysis (Fig 10A). Again, Epo-induced fluorescence was only seen for the 5 and 10 amino acid linked receptor-DHFR fragment fusions, but not for the 30 amino acid linker where the level of fluorescence was independent of ligand. It has been previously demonstrated that the fMTX concentrations in cells directly correlates with the number and activity of DHFR molecules. Because of this, it is possible to calculate the average number of receptors in the cell population, based on the FACS response. Assuming that one EpoR dimer equals one reconstituted DHFR molecule in a 1:1 complexes fMTX (22). An average number of receptors in Epo-activated cells of approximately 11,000 receptors per cell for the 5, 10, and 30 amino acid linker cases was calculated. The fact that the numbers of activated dimers are approximately equal for each construct precludes one obvious problem with this strategy. It might be argued that steric hindrance by other proteins at the membrane intracellular surface might prevent complementation of the DHFR fragments by interfering with simple receptor dimerization in the case of receptors fused to the fragments through short

linkers. However, if this were the case, the constitutive signal seen with the 30 and to some extent with the 10 amino acid linker would be higher than for the activated 5 amino acid linker. In this case the receptors are at a minimum, within the range of 80 Å from each other. In addition, the fact that the number of activated receptors is the same in all cases suggests that no additional, or the same factors, determine the oligomerization state of the receptor for all three cases.

To test whether the ligand induced responses corresponds to the known pharmacological response of EpoR, quantitative FACS analysis of cell fluorescence *versus* ligand concentration (Fig 10B,C) was performed. Both Epo and EMP1 showed saturable binding isotherms with K_ds 164 pM and 168 nM respectively. These values are consistent with previous studies of cellular binding constants and demonstrate that the ligand induced response is consistent with the proposed model. Further, the results are consistent with a single binding constant, typically observed for both Epo and EMP1 binding to receptors expressed on a variety of cell types.

Applicants have shown results consistent with an allosteric mechanism of EpoR activation in the case of the extracellular plus transmembrane domains alone. To demonstrate that this model applies to the complete receptor complex, full length EpoR and JAK2 was coexpressed in COS-7 cells fused to the variable linkers and complementary F[1,2] and F[3] fragments. Results were identical to those observed in CHO cells for the extracellular EpoR domain alone. JAK2 fused to the 5 or 10 linker and F[1,2] or F[3] co-expressed with full length EpoR alone gave an induced response to Epo or EMP1 (Fig 11A). Co-expressed alone, JAK2-5,10-F[1,2] and JAK2-5,10-F[3] showed no constitutive fluorescence, suggesting that they do not interact detectably even when transiently overexpressed. Constitutive fluorescence was seen when JAK2-5,10-F[1,2] was coexpressed with EpoR-5,10-F[3], consistent with previous studies indicating that this interaction is constitutive. However, Epo and EMP1 did induce an augmentation of fluorescence in this case, suggesting that in the activated state, the complex of EpoR with JAK2 is more stable.

Coexpression of EpoR-5aa, 10-F[1,2] with EpoR-5aa or 10aa-F[3] also gave a constitutive response. These results would appear contradictory to the model, but they are not. It was observed, in circular dichroism and NMR studies, that the 236 amino acid intracellular domain of EpoR is not folded. Taken together with the results presented herein, in the fluorescent assay, the intracellular domain acts as a very long linker, resulting in constitutive reconstitution of DHFR. Cotransfected EpoR and JAK2 was also shown to function normally with the attached F[1,2], F[3] fragments. Western blots with anti-EpoR, JAK2 and pTyr show that both proteins are expressed in the cells and that both undergo Epo- or EMP1-induced phosphorylation (Fig. 11).

Based on the results presented here and the structural studies of Wilson, et al, an allosteric model of EpoR activation is proposed. Constitutive dimers of EpoR bring the JAK2 kinases associated with each monomer intracellular domain into contact and allow autophosphorylation and activation of the kinases. This model is not in any way inconsistent with dimerization models; certainly dimerization is required, but not necessarily a sufficient condition for receptor activation. However, given the high sequence and structural homology among the cytokine growth factors it is possible that this model could be generalized to this class of receptors. Furthermore, constitutive dimerization of the cytokine IL-2 receptor, IL-1 receptor and of epidermal growth factor receptors (EGFR) has been detected by quantitative FRET microscopy suggesting that an allosteric mechanism of activation may apply to cytokine and other receptor classes. Simple dimerization, dimerization-allostery or different types of conformation change in constitutive dimers are also possible models for receptor activation. For the structurally and functionally well understood bacterial chemotactic Tar receptors, the mechanism of activation also results from a conformation change in constitutive dimers or tetramers induced by ligand, but the changes are more subtle than those suggested by these results, involving possibly, small piston-, or scissor-like motions or helix supercoiling. The insulin receptor is also known to be a constitutive disulfide cross-linked dimer and might, as a result, not be capabl of the large conformation changes suggested here for EpoR. The DHFR PCA strategy

presented here would be applicable to testing this model in other dimeric or multimeric cytokine receptors or to studying the interactions of other membrane or soluble proteins with activated receptor complexes. The DHFR PCA could also be used in a FRET strategy with proteins fused to fluorescent proteins with complementary absorption-emission spectra, such as mutants of the green fluorescent protein. An important advantage of the DHFR PCA is the absolute requirement that fragments be sufficiently close to fold-reassemble into the enzyme three dimensional structure. This absolute regio- and stereospecific requirement means that a spurious response that might occur between proteins that are merely proximal to each other but not forming a complex is unlikely. Two other advantages include the fact that DHFR is a small, monomeric enzyme. Thus, an observed signal is assured to be due to a dimeric interaction. Second, that the stringency of reassembly can be controlled directly by the introduction of fragment interface mutations that will prevent background reassembly of fragments. However, with sufficient controls, a combination of DHFR and other PCAs used in a FRET strategy would provide a powerful approach to studies of protein association dynamics throughout the cell in localized regions and compartments.

Of course, numerous membrane receptors can be used when screening for agonist and antagonist. The receptors of interest include the Erythropoietin receptor as well as the following additional cellular receptors: receptor from a member of a protein family selected from the group consisting of the TGF-beta, NGF, FGF/HBGF, chemokine, IL-6, LIF/OSM, TNF, MDK/PTN families, Mullerian inhibitory substance (MIS), the inhibins (INHA and INHB), the bone morphogenic proteins (BMP), the growth development factors (GDF-1, GDF-3, GDF-5, GDF-6, GDF-7 and GDF-8), endometrial bleeding associated factor (EBAF/Lefty), glial cell line-derived neurotrophic factor (GDNF), nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3), NT-4 and NT-5, fibroblast growth factor-3 (FGF-3), FGF-4 (int-2), FGF-5, FGF-6 (hst-2), keratinocyte growth factor (KGF/FGF-7), androgen-induced growth factor (AIGF/FGF-8), glia-activating factor (GAF/FGF-9), FGF-11, FGF-12, FGF-13, and FGF-14, platelet factor 4 (PF4), platelet basic protein (PBP),

monocyte-derived neutrophil chemotactic factor (MDNCF/IL-8), melanoma growth stimulatory activity protein (MGSA), macrophage inflammatory protein 2 (MIP-2), Mig, chicken 9E3, pig aveolar macrophage chemotactic factor, pre-B cell growth stimulatory factor (PBSF), cytokine-induced neutrophil chemoattractant-2, IP10, monocyte chemotactic protein 1, (MCP-1), MCP-2, MCP-3, MCP-4, MCP-5, MIP-1-alpha, MIP-1-beta, MIP-1-gamma, MIP-3-alpha, MIP-3-beta, MIP-4, MIP-5, RANTES, SIS-epsilon, thymus and activation-regulated chemokine (TARC), eotaxin, I-309, HCC-1/NCC-2, HCC-3, 6Ckine/Exodus-2/SLC, thymus -expressed chemokine (TECK), mouse protein C10, IL-6, granulocyte colony-stimulating factor (G-CSF), and myelomonocytic growth factor (MGF), leukemia inhibitory factor (LIF) and oncostatin (OSM), tumor necrosis factor alpha (TNF-a), tumor necrosis factor beta (TNF-b/LT-a), CD40L, CD137L/4-1BBL, CD134L/OX40L, CD27L/CD70, FasL, CD30L, LT-b, TNF-related apoptosis-inducing ligand (TRAIL), macrophage stimulating protein, hepatocyte growth factor, platelet-derived growth factor, insulin-like growth factor, platelet-derived endothelial cell growth factor, IL-1a, IL-1b, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11,IL-12, IL-13, IL-14, IL-15, IL-16, IL-17 and IL-18. Other receptors include nuclear receptors or coactivators such as the vitamin D receptors, retinoid receptors, steroid receptors and gamma PPAR receptors.

As shown above, the instant invention allows:

1) the detection of protein-protein interactions *in vivo* or *in vitro*.

2) the detection of protein-protein interactions in appropriate contexts, such as within a specific organism, cell type, cellular compartment, or organelle.

3) the detection of induced versus constitutive protein-protein interactions (such as by a cell growth or inhibitory factor).

4) to distinguish specific-versus non-specific protein-protein interactions by controlling the sensitivity of the assay.

5) the detection of the kinetics of protein assembly in cells.

6) screening of cDNA libraries for protein-protein interactions.

Further aspects of the invention can be demonstrated by identifying novel interactions with the enzyme p70S6k, to determine its' regulation and how separate signaling cascades converge on this enzyme.

The PCA method is particularly useful for detection of the kinetics of protein assembly in cells. The kinetics of protein assembly can be determined using fluorescent protein systems.

In a further embodiment of the invention, PCA can be used for drug screening. The techniques of PCA are used to screen for drugs that block specific biochemical pathways in cells allowing for a carefully targeted and controlled method for identifying products that have useful pharmacological properties.

Although the present invention has been described herein above by way of preferred embodiments thereof, it can be modified, without departing from the spirit and nature of the subject invention as defined in the appended claims.

REFERENCES

1. Reed, L.J., 1974, Acc. Chem. Res. 7:40-46.
2. Lander, E.S. 1996, Science 274:536-539.
3. Evangelista et al. 1996, Trends in Cell Biology 6:196-199.
4. Guarente, L. 1993, Proc. Natl. Acad. Sci. USA 90:1639-41.
5. Adams et al. 1991, Nature 349:694-7.
6. Chien et al. 1991, Proc. Natl. Acad. Sci. USA 88:9578-82.
7. Fields et al. 1989, Nature 340:245-6.
8. Gyuris et al. 1993, Cell 75:791-803.
9. Johnsson et al. 1994, Proc. Natl. Acad. Sci. USA 91:10340-4.
10. Volz et al. 1982, J. Biol. Chem. 257:2528-36.
11. Oefner et al. 1988, Eur. J. Biochem. 174:377-85.
12. Filman et al. 1982, J. Biol. Chem. 257:13663-72.
13. Bystroff et al. 1991, Biochemistry 30:2227-39.
14. Appleman et al. 1988, J. Biol. Chem. 263:10304-13.
15. Loetscher et al. 1991, J. Biol. Chem. 266:11213-20.
16. Kaufman et al. 1989, Mol. Cell. Biol. 9:946-58.
17. Stammers et al. 1987, FEBS Lett. 218:178-84.
18. Gegg et al., Techniques in Protein Chemistry (eds. Marshak, D.R.) 439-448 (Academic Press, New York, USA, 1996).
19. Perry et al. 1989, Biochemistry 28:7961-8.
20. Bullerjahn et al. 1992, J. Biol. Chem. 267:864-70.
21. Buchwalder et al. 1992, Biochemistry 31:1621-30.
22. Chen et al. 1992, Biochemistry 31:2219-23.
23. Prevost et al. 1991, Proc. Natl. Acad. Sci. USA 88:10880-4.
24. Kellis et al. 1988, Nature 333:784-6.
25. Kellis et al. 1989, Biochemistry 28:4914-22.
26. Henderson et al., 1980, Arch. Biochem. Biophys. 202:29-34.
27. Denzer et al. 1995, EMBO J. 14:6311-7.
28. Spiess et al. 1985, J. Biol. Chem. 260:1979-82.
29. Picard et al. 1994, Nucleic Acids Res. 22:2587-91.

30. Kunkel et al. 1987, Methods Enzymol. 154:367-82.
31. Tartof et al. 1988, Gene 67:169-82.
32. Laemmli, U.K. 1970, Nature 227:680-5.
33. Ellenberger et al. 1992, Cell 71:1223-37.
34. Cody et al. 1992, Anti Cancer Drug Des. 7:483-91.
101. Reed, L.J. 1974, Acc. Chem. Res. 7:40-46.
102. Guarente, L. 1993, Proc. Natl. Acad. Sci. USA 901639-41 (1993).
103. Adams et al. 1991, Nature 349:694-7.
104. Chien et al. 1991, Proc. Natl. Acad. Sci. USA 88:9578-82.
105. Fields et al. 1989, Nature 340:245-6.
106. Gyuris et al. 1993, Cell 75:791-803.
107. Guarente, L. 1993, Proc. Natl. Acad. Sci. USA 90:1639-41.
108. Johnsson et al. 1994, Proc. Natl. Acad. Sci. USA 91:10340-4.
109. O'Shea et al. 1991, Science 254:539-44.
110. Ellenberger et al. 1992, Cell 71:1223-37.
111. Volz et al. 1982, J. Biol. Chem. 257:2528-36.
112. Oefner et al. 1988, Eur. J. Biochem. 174:377-85.
113. Filman et al. 1982, J. Biol. Chem. 257:13663-72.
114. Bystroff et al. 1991, Biochemistry 30:2227-39.
115. Jones et al. 1995, Protein Sci. 4:167-77.
116. Jennings et al. 1993, Biochemistry 32:3783-9.
117. Fierke et al. 1987, Biochemistry 26:4085-92.
118. Andrews et al. 1989, Biochemistry 28:5743-50.
119. Thillet et al. 1990, Biochemistry 29:5195-202.
120. Hillcoat et al. 1967, Anal. Biochem. 21:178-89.
121. Appleman et al. 1988, J. Biol. Chem. 263:10304-13.
122. Grange et al. 1984, Nucleic Acids Res. 12:3585-601.
123. Loetscher et al. 1991, J: Biol. Chem. 266:11213-20.
124. Hao et al. 1994, J. Biol. Chem. 269:15179-85.
125. Kaufman, R.J. 1985, Proc. Natl. Acad. Sci. USA 82:689-93.
126. Kaufman et al. 1989, Mol. Cell. Biol. 9:946-58.
127. Oefner et al. 1988, Eur. J. Biochem. 174:377-85.

128. Stammers et al. 1987, FEBS Lett. 218:178-84.

129. Gegg et al., Techniques in Protein Chemistry (eds. Marshak, D.R.) 439-448 (Academic Press, New York, USA, 1996).

130. Bystroff et al. 1991, Biochemistry 30:2227-39.

131. Harmann et al. 1990, FEBS Lett. 273:248-52.

132. Kozma et al. 1990, Proc. Natl. Acad. Sci. USA 87:7365-9.

133. Grove et al. 1991, Mol. Cell. Biol. 11:5541-50.

134. Banerjee et al. 1990, Proc. Natl. Acad. Sci. USA 87:8550-4.

135. Reinhard et al. 1992, Proc. Natl. Acad. Sci. USA 89:4052-6.

136. Reinhard et al. 1994, EMBO J. 13:1557-65.

137. Chung et al. 1992, Cell 69:1227-36.

138. Price et al. 1992, Science 257:973-7.

139. Kuo et al. 1992, Nature 358:70-3.

140. Terada et al. 1993, J. Biol. Chem. 268:12062-8.

141. Calvo et al. 1992, Proc. Natl. Acad. Sci. USA 89:7571-5.

142. Weng et al. 1995, Mol. Cell. Biol. 15:2333-40.

143. Ming et al. 1994, Nature 371:426-9.

144. Cheatham et al. 1995, Proc. Natl. Acad. Sci. USA 92:11696-700.

145. Flotow et al. 1992, J. Biol. Chem. 267:3074-8.

146. Wendt et al. 1994, J. Am. Chem. Soc. 116:6973-6974.

147. Mahalingam et al. 1996, Mol. Cell. Biol. 16:405-13.

148. Edery et al. 1995, Mol. Cell. Biol. 15:3363-71.

149. Hussain et al. 1992, Gene 112:179-88.

150. Lim et al. 1994, Protein Sci. 3:2233-44.

151. Habig et al. 1975, Biochemical & Biophysical Research Communications 64:501-6.

152. Chalfie et al. 1994, Science 263:802-5.

153. Cody et al. 1993, Biochemistry 32:1212-8.

154. Morin et al. 1971, Journal of Cellular Physiology 77:313-8.

155. Morise et al. 1974, Biochemistry 13:2656-62.

156. Ward et al. 1982, Biochemistry 21:4535-40.

157. Heim et al. 1994, Proc. Natl. Acad. Sci. USA 91:12501-4.

158. Ormo et al. 1996, Science 273:1392-1395.

159. Youvan et al. 1996, Nature Biotechnology 14:1219-1220.

160. DeLuca et al. 1986, Analytical Biochemestry 161:501-507.

161. Conti et al. 1996, Structure 4:287-298.

162. Magrath et al. 1993, Biotechniques 15:4-6.

Abbreviations: PCA, Protein-fragment Complementation Assay; mDHFR, murine dihydrofolate reductase; hDHFR, human dihydrofolate reductase; Z-F[1,2], GCN4 leucine zipper-mDHFR fragment[1,2]; USPS, ubiquitiN-based split-protein sensor; IPTG, isopropyl-b-D-thiogalactopyranoside; PMSF, phenylmethylsulfonyl fluoride; SDS-PAGE, SDS polyacrylamide gel electrophoresis.

WHAT IS CLAIMED IS:

1. A method employing a Protein Complementation assay/Universal Reporter System (PCA/URS) for detecting and screening for agonists and antagonists of a cellular receptor, which method comprises:

a) generating a first nucleic acid vector encoding a first fusion product comprising:

i) a first fragment of a first PCA/URS reporter molecule, and

ii) a second molecule, fused to said first fragment, which comprises a first subdomain of a cellular receptor molecule of interest;

b) generating a second nucleic acid vector encoding a second fusion product comprising:

i) a second fragment of said first PCA/URS reporter molecule, and

ii) a third molecule, fused to said second fragment, which comprises a second subdomain of said cellular receptor, and where said second subdomain may be the same as said first subdomain in the case of a homodimeric cellular receptor, or different from said first subdomain in the case of a heterodimeric cellular receptor; or a receptor coactivator or a protein;

c) transfecting prokaryotic or eukaryotic cells with said first and second nucleic acid vectors;

d) testing said transfected cells for the PCA/URS reporter activity, said activity indicating reassociation of the first and second fragments of the PCA/URS molecule mediated by the interaction of said first and second subdomains of the cellular receptor molecule; said association being induced by binding said receptor to cognate ligand.

2. A method employing a Protein Complementation Assay/Universal Reporter System (PCA/URS) for detecting and screening for agonists and antagonists of a cellular receptor, which method comprises:

a) generating a first nucleic acid vector encoding a first fusion product comprising:

i) a first fragment of a first PCA/URS reporter molecule, and

ii) a second molecule, fused to said first fragment, which comprises a first subdomain of a cellular receptor molecule of interest;

b) generating a second nucleic acid vector encoding a second fusion product comprising:

i) a second fragment of said first PCA/URS reporter molecule, and

ii) a third molecule, fused to said second fragment, which comprises a second subdomain of said cellular receptor, and where said second subdomain may be the same as said first subdomain in the case of a homodimeric cellular receptor, or different from said first subdomain in the case of a heterodimeric cellular receptor;

c) transfecting prokaryotic or eukaryotic cells with said first and second nucleic acid vectors;

d) obtaining a clonal population of cells that express said first and second fusion products; and

e) testing said transfected cells for the PCA/URS reporter activity, said activity indicating reassociation of the first and second fragments of the PCA/URS molecule mediated by the interaction of said first and second subdomains of the cellular receptor molecule; said association being induced by binding said receptor to cognate ligand.

3. The method of claim 2, further comprising the step of treating said clonal population of cells with a chemical composition prior to said

testing of the cells for PCA/URS activity, thus measuring the ability of the chemical composition to induce or inhibit the activity.

4. The method of claim 3, wherein said chemical composition is an individual compound or a mixture of compounds obtained from a chemical compound library or combinatorial chemical synthesis.

5. The method of claim 2, wherein said reporter molecule is a multimeric protein.

6. The method of claim 2, wherein said reporter molecule is a multimeric receptor.

7. The method of claim 2, wherein said reporter molecule is a multimeric binding protein.

8. The method of claim 2, wherein said reporter molecule is a catalytic molecule.

9. The method of claim 2, wherein said reporter molecule is an energy transfer molecule.

10. The method of claim 3, wherein said reporter molecule is a multimeric protein.

11. The method of claim 2, wherein said reporter molecule is a fluorescent, luminescent or phosphorescent protein.

12. The method of claim 2, wherein said reporter molecule is an electron transfer molecule.

13. The method of claim 2, wherein said reporter molecule is a chemiluminescent molecule.

14. The method of claim 3, wherein said chemical composition is a ligand agonist or antagonist.

15. The method of claim 3, wherein said chemical composition is a nucleic acid.

16. The method of claim 3, wherein said chemical composition is a peptide.

17. The method of claim 3, wherein said chemical composition is a carbohydrate.

18. The method of claim 3, wherein said chemical composition is a natural product or extract.

19. The method of claim 4, wherein said library of compounds is a combinatorial nucleic acid library.

20. The method of claim 4, wherein said library of compounds is a combinatorial carbohydrate library.

21. The method of claim 4, wherein said library of compounds is a combinatorial peptide or protein library.

22. The method of claim 3, wherein in the treatment step the cells are treated with the chemical composition at different concentrations in the medium, and the PCA/URS activity is compared at the different concentrations.

23. The method of claim 22, wherein the values of PCA/URS activity versus concentration of treatment composition are used to estimate the binding isotherm of the composition to the cellular receptor.

24. The method of claim 2, wherein the FCA/URS activity is detected using a fluorescent assay, and the activity is monitored by fluorescence microscopy, fluorescent cell sorting (FACS) or by spectroscopy of aliquots of the cells.

25. The method of claim 22, wherein said reporter molecule is dihydrofolate reductase and said detection method comprises treatment of the cells with fluorescein-conjugated methotrexate before monitoring the cellular fluorescence.

26. The method of claim 2, wherein said cellular receptor is the Erythropoietin receptor.

27. The method of claim 2, wherein said cellular receptor is a naturally occuring protein which upon binding a ligand induces a cellular response.

28. The method of claim 2, wherein said cellular receptor is an enzyme which is activated by binding a ligand.

29. The method of claim 2, wherein said cellular receptor is a natural or synthetic protein which undergoes conformational change or oligomerizes upon binding a ligand.

30. The method of claim 2, wherein said cellular receptor is a member of the cytokine receptor superfamily.

31. The method of claim 2, wherein said cellular receptor is the receptor for an interleukin or cytokine.

32. The method of claim 2, wherein said cellular receptor is a hormone receptor.

33. The method of claim 2, wherein said cellular receptor is a receptor for a member of a protein family selected from the group consisting of the TGF-beta, NGF, FGF/HBGF, chemokine, IL-6, LIF/OSM, TNF, and MDK/PTN families.

34. The method of claim 2, wherein said cellular receptor is the receptor for a member of the tumor growth factor beta family.

35. The method of claim 2, wherein said cellular receptor is the receptor for a protein selected from the group consisting of the forms of TGF-beta, Mullerian inhibitory substance (MIS), the inhibins (INHA and INHB), the bone morphogenic proteins (BMP), the growth development factors (GDF-1, GDF-3, GDF-5, GDF-6, GDF-7 and GDF-8), endometrial bleeding associated factor (EBAF/Lefty), and glial cell line-derived neurotrophic factor (GDNF).

36. The method of claim 2, wherein said cellular receptor is the receptor for a member of the nerve growth factor family.

37. The method of claim 2, wherein said cellular receptor is the receptor for a protein selected from the group consisting of nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3), NT-4 and NT-5.

38. The method of claim 2, wherein said cellular receptor is the receptor for a member of the fibroblast growth factor and heparin-binding growth factor family.

39. The method of claim 2, wherein said cellular receptor is the receptor for a protein selected from the group consisting of fibroblast growth factor-3 (FGF-3), FGF-4 (int-2), FGF-5, FGF-6 (hst-2), keratinocyte growth factor (KGF/FGF-7), androgen-induced growth factor (AIGF/FGF-8), glia-activating factor (GAF/FGF-9), FGF-11, FGF-12, FGF-13, and FGF-14.

40. The method of claim 2, wherein said cellular receptor is the receptor is the receptor for a member of the chemokine family.

41. The method of claim 2, wherein said cellular receptor is the receptor for a protein selected from the group consisting of platelet factor 4 (PF4), platelet basic protein (PBP), monocyte-derived neutrophil chemotactic factor (MDNCF/IL-8), melanoma growth stimulatory activity protein (MGSA), macrophage inflammatory protein 2 (MIP-2), Mig, chicken 9E3, pig aveolar macrophage chemotactic factor, pre-B cell growth stimulatory factor (PBSF), cytokine-induced neutrophil chemoattractant-2, and IP10.

42. The method of claim 2, wherein said cellular receptor is the receptor for a protein selected from the group consisting of monocyte chemotactic protein 1, (MCP-1), MCP-2, MCP-3, MCP-4, MCP-5, MIP-1-alpha, MIP-1-beta, MIP-1-gamma, MIP-3-alpha, MIP-3-beta, MIP-4, MIP-5, RANTES, SIS-epsilon, thymus and activation-regulated chemokine (TARC), eotaxin, I-309, HCC-1/NCC-2, HCC-3, 6Ckine/Exodus-2/SLC, thymus -expressed chemokine (TECK) and mouse protein C10.

43. The method of claim 2, wherein said cellular receptor is the receptor for a protein selected from the group consisting of fractalkine and GCP-2/LIX.

44. The method of claim 2, wherein said cellular receptor is a member of the group consisting of CXCR-1, CXCR-2, CXCR-3, CXCR-4, CCR-1, CCR-2, CCR-3, CCR-4, CCR-5, CCR-6, CCR-7, CCR-8, and CX3CR.

45. The method of claim 2, wherein said cellular receptor is the receptor for a member of the interleukin-6 (IL-6) family.

46. The method of claim 2, wherein said cellular receptor is the receptor for a protein selected from the group consisting of IL-6, granulocyte colony-stimulating factor (G-CSF), and myelomonocytic growth factor (MGF).

47. The method of claim 2, wherein said cellular receptor is the receptor for a member of the leukemia inhibitory factor and oncostatin family.

48. The method of claim 2, wherein said cellular receptor is the receptor for a member of the group selected from leukemia inhibitory factor (LIF) and oncostatin (OSM).

49. The method of claim 2, wherein said cellular receptor is the receptor for a member of the tumor necrosis factor family.

50. The method of claim 2, wherein said cellular receptor is the receptor for a protein selected from the group consisting of tumor necrosis factor alpha (TNF-a), tumor necrosis factor beta (TNF-b/LT-a), CD40L, CD137L/4-1BBL, CD134L/OX40L, CD27L/CD70, FasL, CD30L, LT-b and TNF-related apoptosis-inducing ligand (TRAIL).

51. The method of claim 2, wherein said cellular receptor is a receptor selected from the group consisting of LNGFR/p75, CD40, CD137/4-1BB/ILA, TNFRI/p55/CD120a, TNFRII/p75/CD120b, CD134/OX40/ACT35, CD27, Fas/CD95/APO-1, CD30/Ki-1, LT-betaR, DR3/WSL-1/TRAMP/APO-3/LARD, DR4, DR5, DcR1/TRID, TR2, GITR and osteoprotegerin (OPG).

52. The method of claim 2, wherein said cellular receptor is the receptor for a member of the midkine and pleiotrophin family.

53. The method of claim 2, wherein said cellular receptor is the receptor for a protein selected from the group consisting of midkine (MK), pleiotrophin (PTN), chicken retinoic acid-induced heparin-binding protein (RI-MB), Xenopus pleiotrophic factors alpha-1, alpha-2, beta-1 and beta-2.

54. The method of claim 2, wherein said cellular receptor is a member of the family of G-protein-coupled receptors.

55. The method of claim 2, wherein said cellular receptor is a receptor for transferrin.

56. The method of claim 2, wherein said cellular receptor is a receptor for a member of the group consisting of macrophage stimulating protein, hepatocyte growth factor, platelet-derived growth factor, insulin-like growth factor and platelet-derived endothelial cell growth factor.

57. The method of claim 2, wherein said cellular receptor is the receptor for a steroid hormone.

58. The method of claim 2, wherein said cellular receptor is the receptor for an eicosanoid hormone.

59. The method of claim 2, wherein said cellular receptor has been identified from an expressed sequence tag (EST) nucleic acid sequence.

60. The method of claim 2, wherein said cellular receptor is the receptor for a protein selected from the group consisting of IL-1a, IL-1b, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11,IL-12, IL-13, IL-14, IL-15, IL-16, IL-17 and IL-18.

61. The method of claim 2 wherein said receptor is a nuclear receptor or coativator of said nuclear receptor.

62. The method of claim 61 wherein said receptor is the Vitamin D receptor.

63. The method of claim 61 wherein said receptor is a Vitamin A or a retinoid associated receptor.

64. The method of claim 61 wherein said receptor is a Gamma PPAR.

65. The method of claim 61 wherein said receptor is a steroid receptors.

66. A method employing a Protein Complementation Assay/Universal Reporter System (PCA/URS) for detecting and screening for agonists and antagonists of a membrane receptor, which method comprises:

a) generating a first nucleic acid vector encoding a first fusion product comprising:

i) a first fragment of a first PCA/URS reporter molecule,

ii) a first linker, fused at one end to said first fragment, said linker region comprising between 1 and 30 amino acid residues; and

iii) a second molecule, fused to the other end of said first linker, which comprises a first subdomain of a cellular receptor molecule of interest;

b) generating a second nucleic acid vector encoding a second fusion product comprising:

i) a second fragment of said first PCA/URS reporter molecule,

ii) a second linker, fused at one end to said second fragment, said linker comprising between 1 and 30 amino acid residue; and

iii) a third molecule, fused to the other end of said second linker,

which comprises a second subdomain of said cellular receptor, and where said second subdomain may be the same as said first subdomain in the case of a homodimeric cellular receptor, or different from said first subdomain in the case of a heterodimeric cellular receptor;

c) transfecting prokaryotic or eukaryotic cells with said first and second nucleic acid vectors;

d) testing said transfected cells for the PCA/URS reporter activity, said activity indicating reassociation of the first and second fragments of the PCA/URS molecule mediated by the interaction of said first and second subdomains of the cellular receptor molecule.

Figure 1

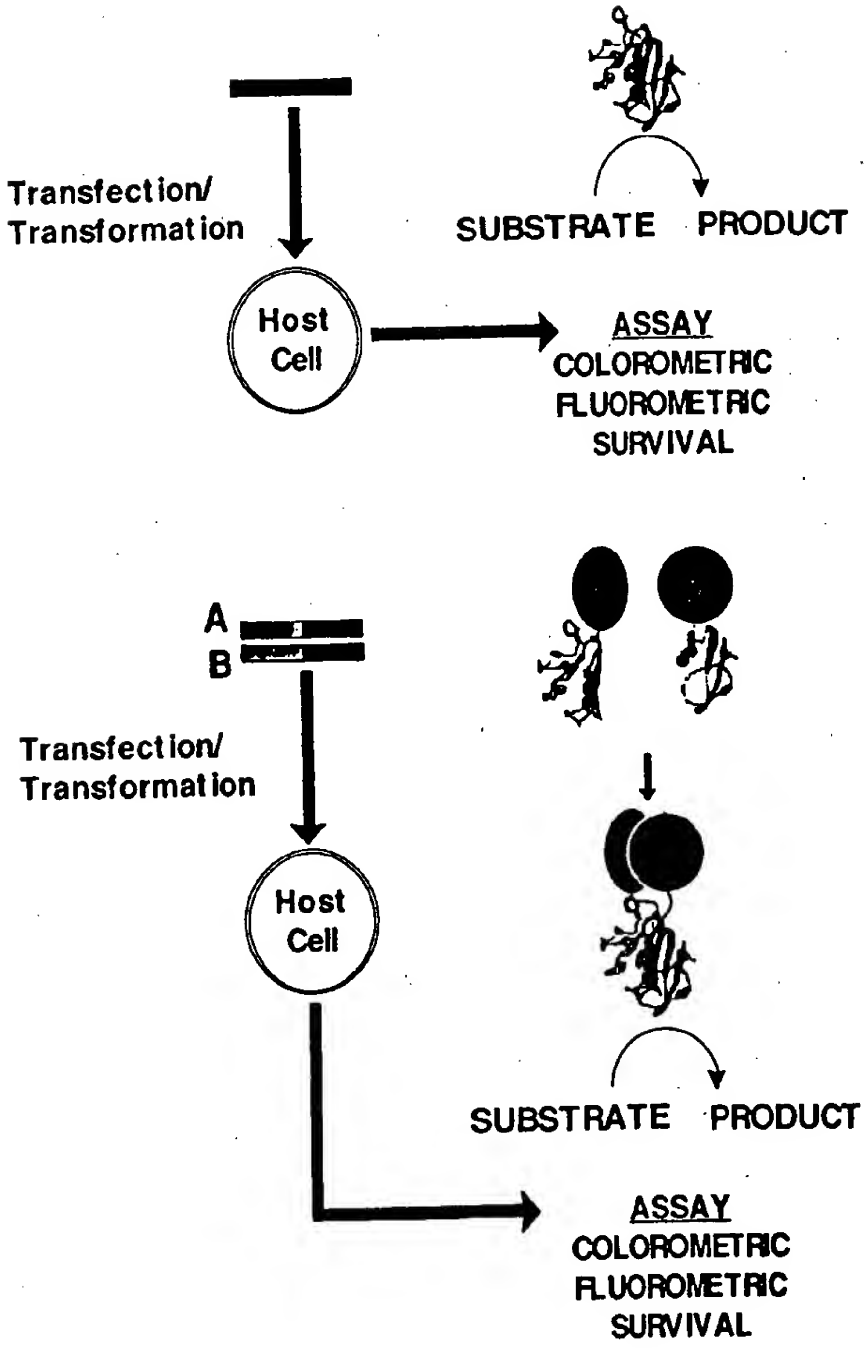

Transfection/
Transformation
Host
Cell
SUBSTRATE
PRODUCT
ASSAY
COLOROMETRIC
FLUOROMETRIC
SURVIVAL
A
B
Transfection/
Transformation
Host
Cell
SUBSTRATE
PRODUCT
ASSAY
COLOROMETRIC
FLUOROMETRIC
SURVIVAL

Figure 2

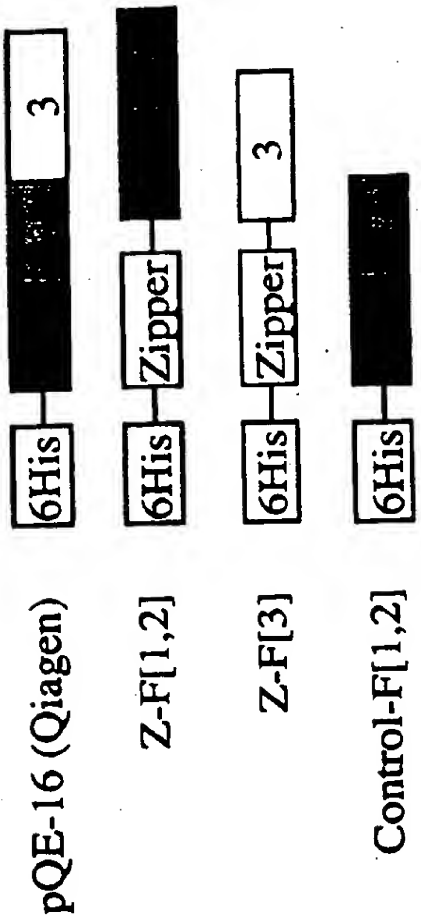

pQE-16 (Qiagen)
6His
3
Z-F[1,2]
6His
Zipper
Z-F[3]
6His
Zipper
3
Control-F[1,2]
6His

Figure 3

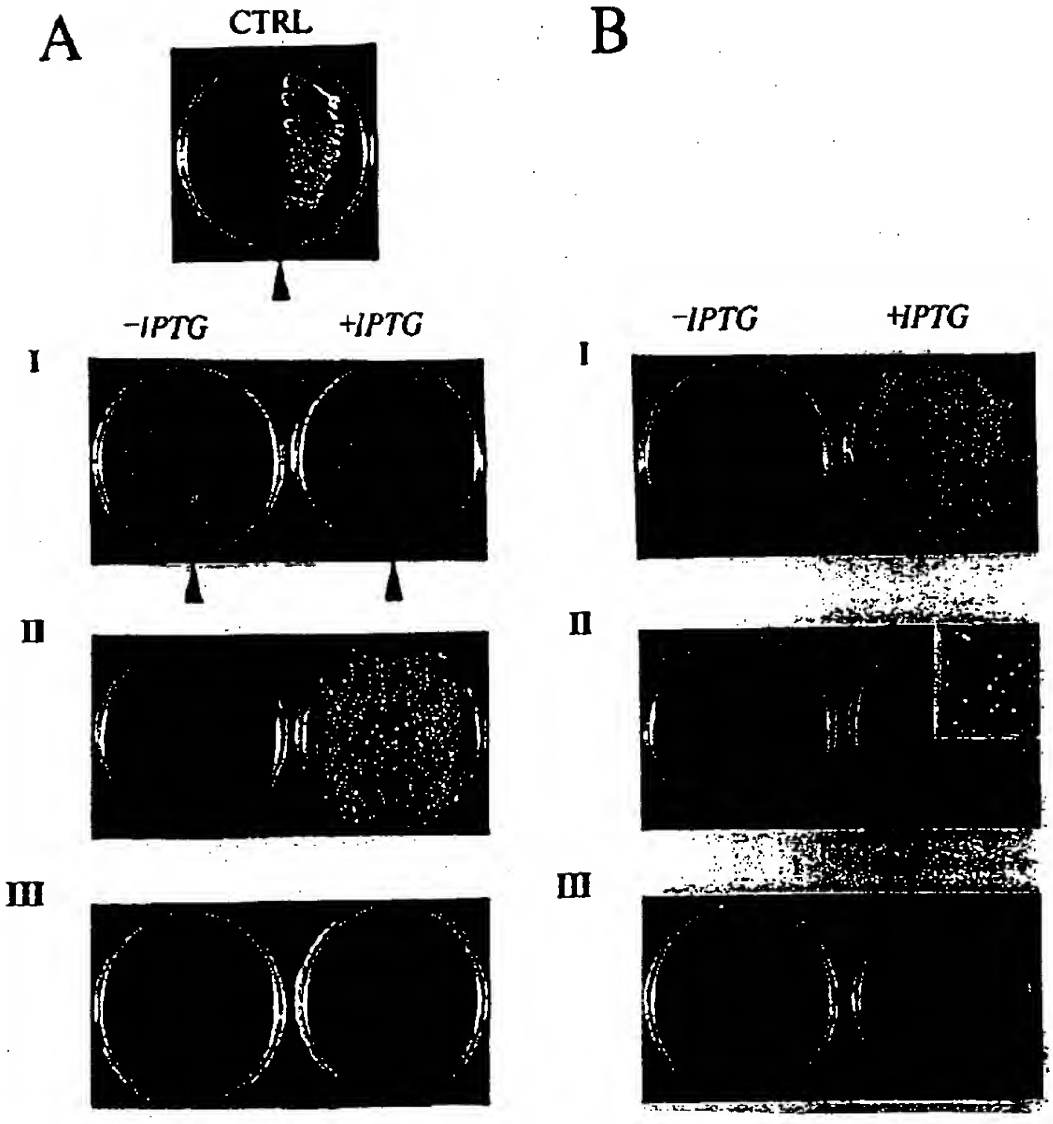


Figure 4

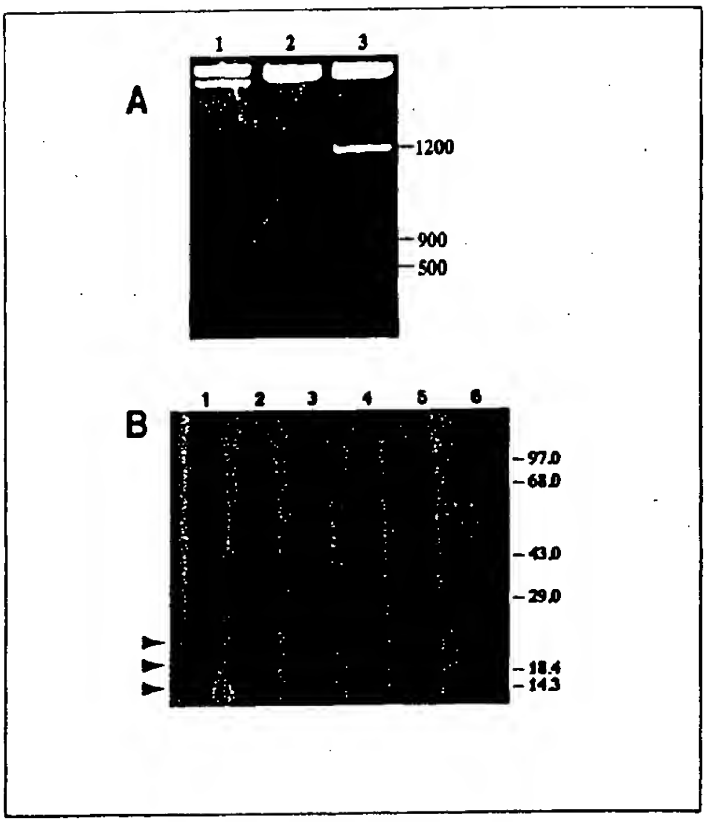


Figure 5

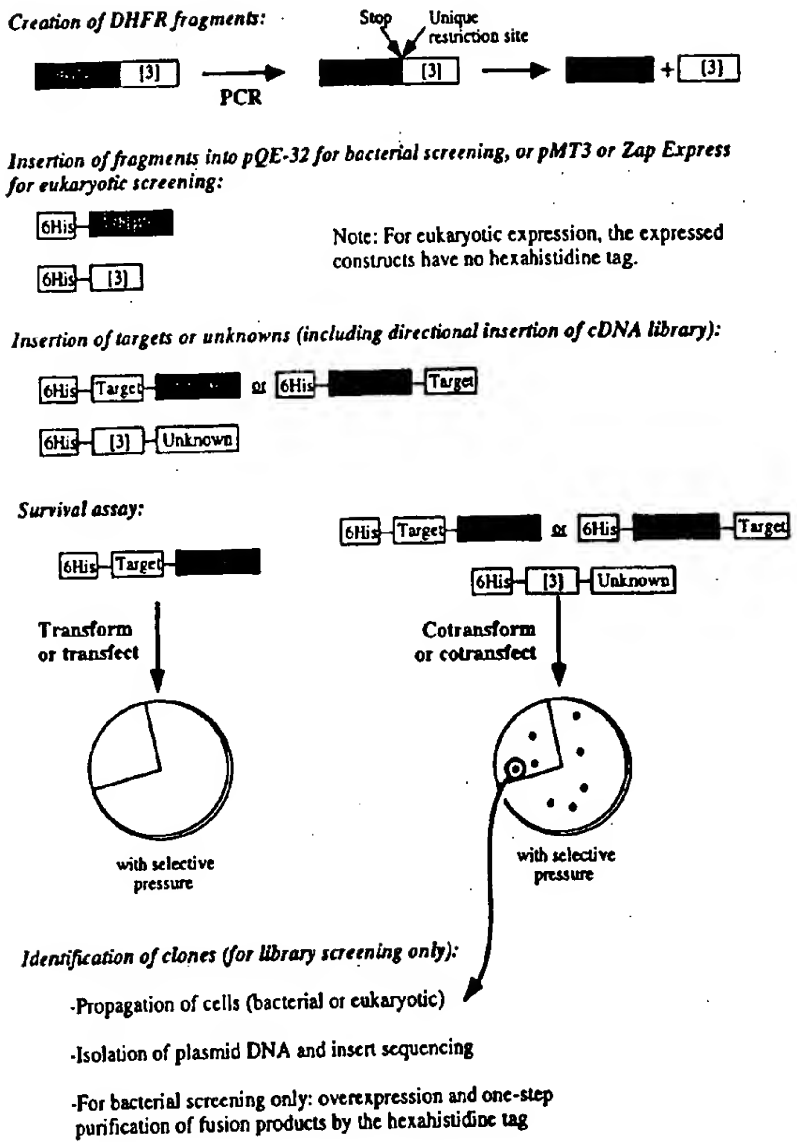

Creation of DHFR fragments:
Stop
Unique restriction site
[3]
PCR
[3]
+
[3]
Insertion of fragments into pQE-32 for bacterial screening, or pMT3 or Zap Express for eukaryotic screening:
6His
6His
[3]
Note: For eukaryotic expression, the expressed constructs have no hexahistidine tag.
Insertion of targets or unknowns (including directional insertion of cDNA library):
6His
Target
or
6His
Target
6His
[3]
Unknown
Survival assay:
6His
Target
6His
Target
or
6His
Target
6His
[3]
Unknown
Transform or transfect
Cotransform or cotransfect
with selective pressure
with selective pressure
Identification of clones (for library screening only):
-Propagation of cells (bacterial or eukaryotic)
-Isolation of plasmid DNA and insert sequencing
-For bacterial screening only: overexpression and one-step purification of fusion products by the hexahistidine tag

Figure 6

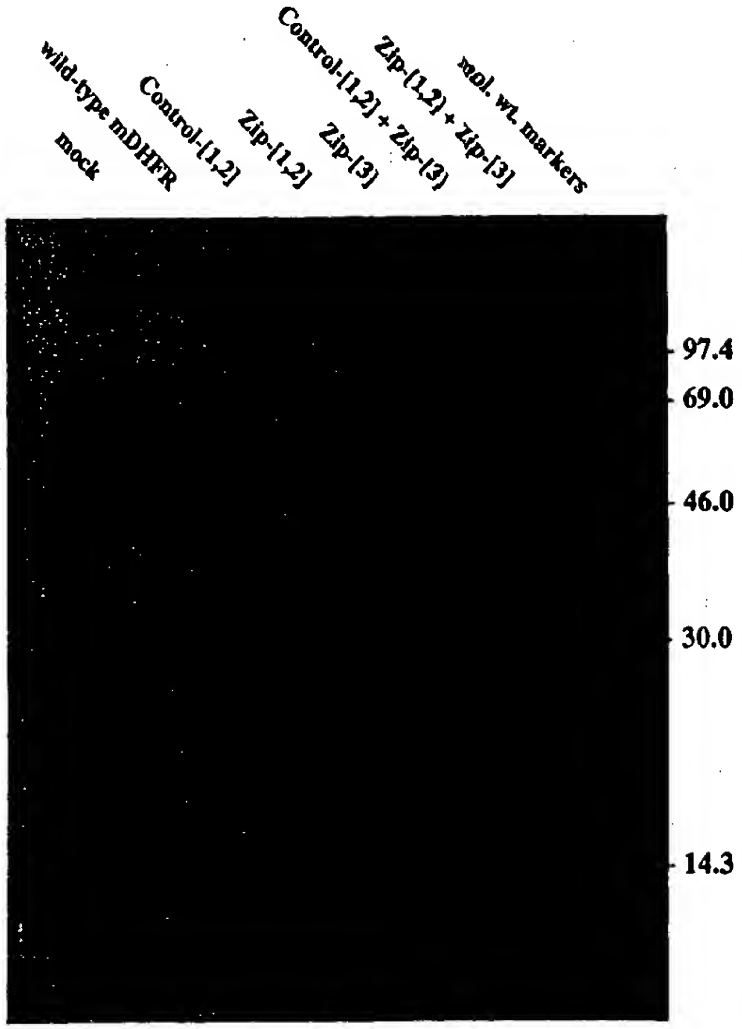
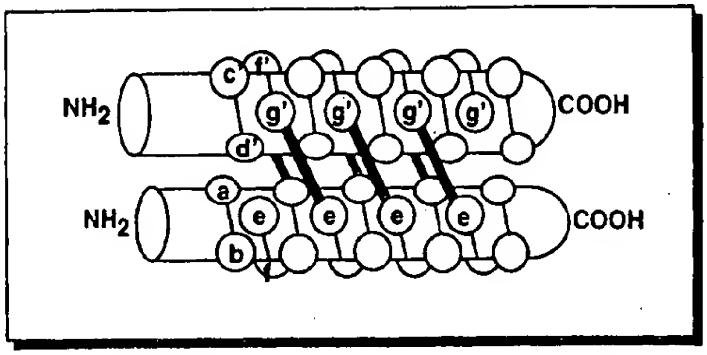

mock
wild-type mDHFR
Control-[1,2]
Zip-[1,2]
Zip-[3]
Control-[1,2] + Zip-[3]
Zip-[1,2] + Zip-[3]
mol. wt. markers
97.4
69.0
46.0
30.0
14.3

Figure 7

Distribution of e-g Pairs: Zipper Library-1 vs Zipper Library-2 Screen



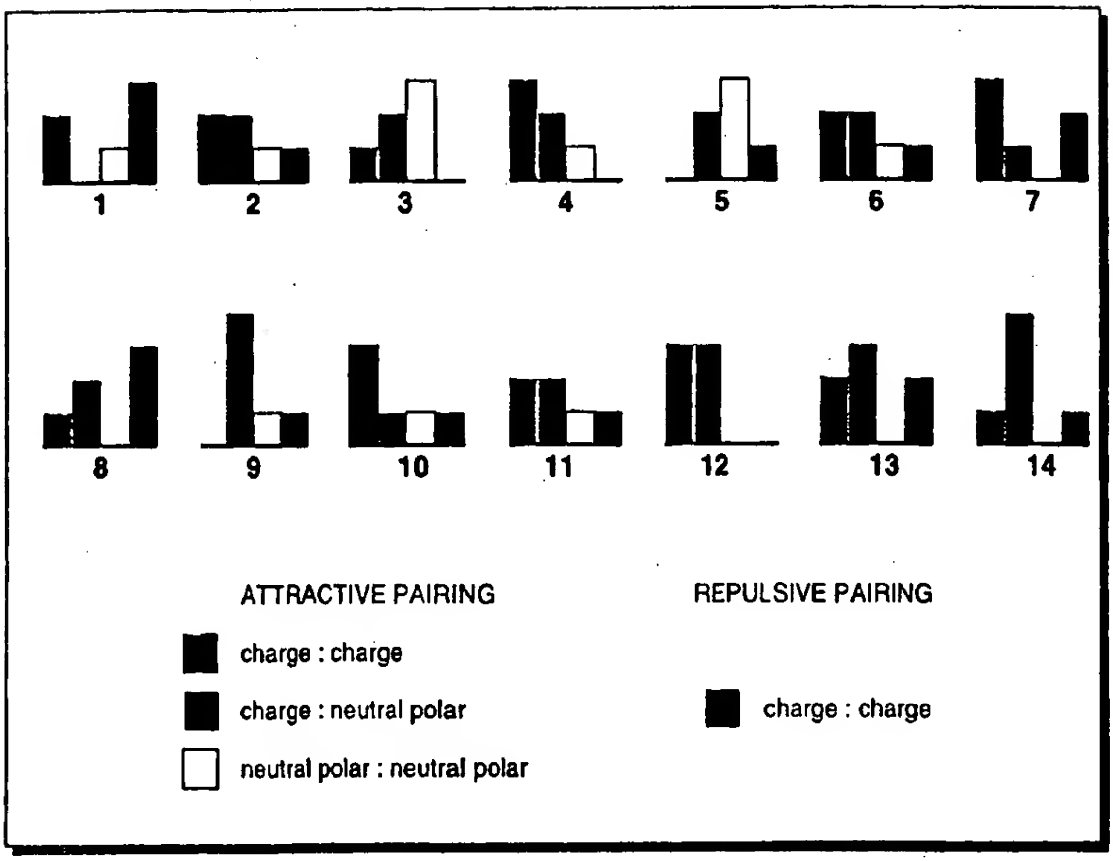


FIGURE 8

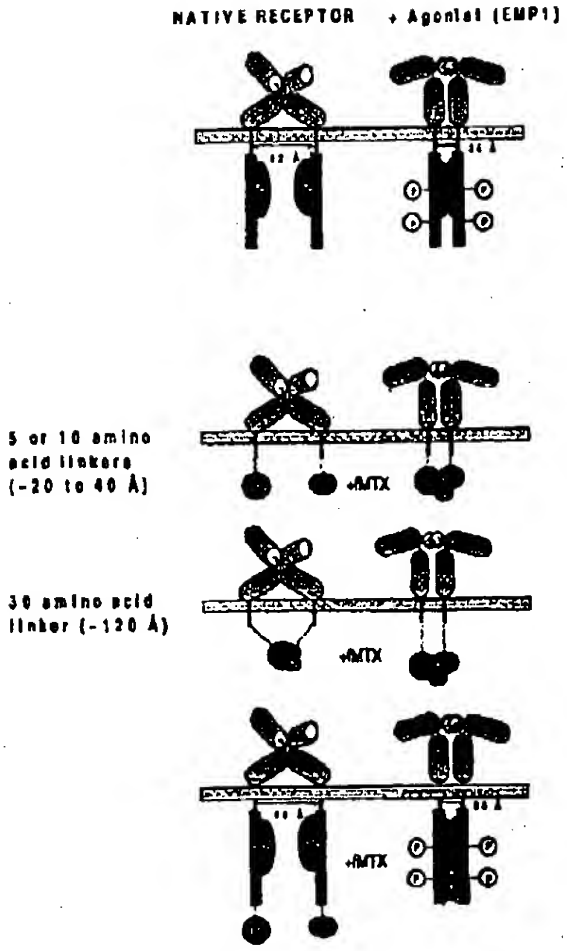

NATIVE RECEPTOR + Agonist (EMP1)
5 or 10 amino acid linkers (~20 to 40 Å)
+MTX
30 amino acid linker (~120 Å)
+MTX
+MTX

Fig. 9

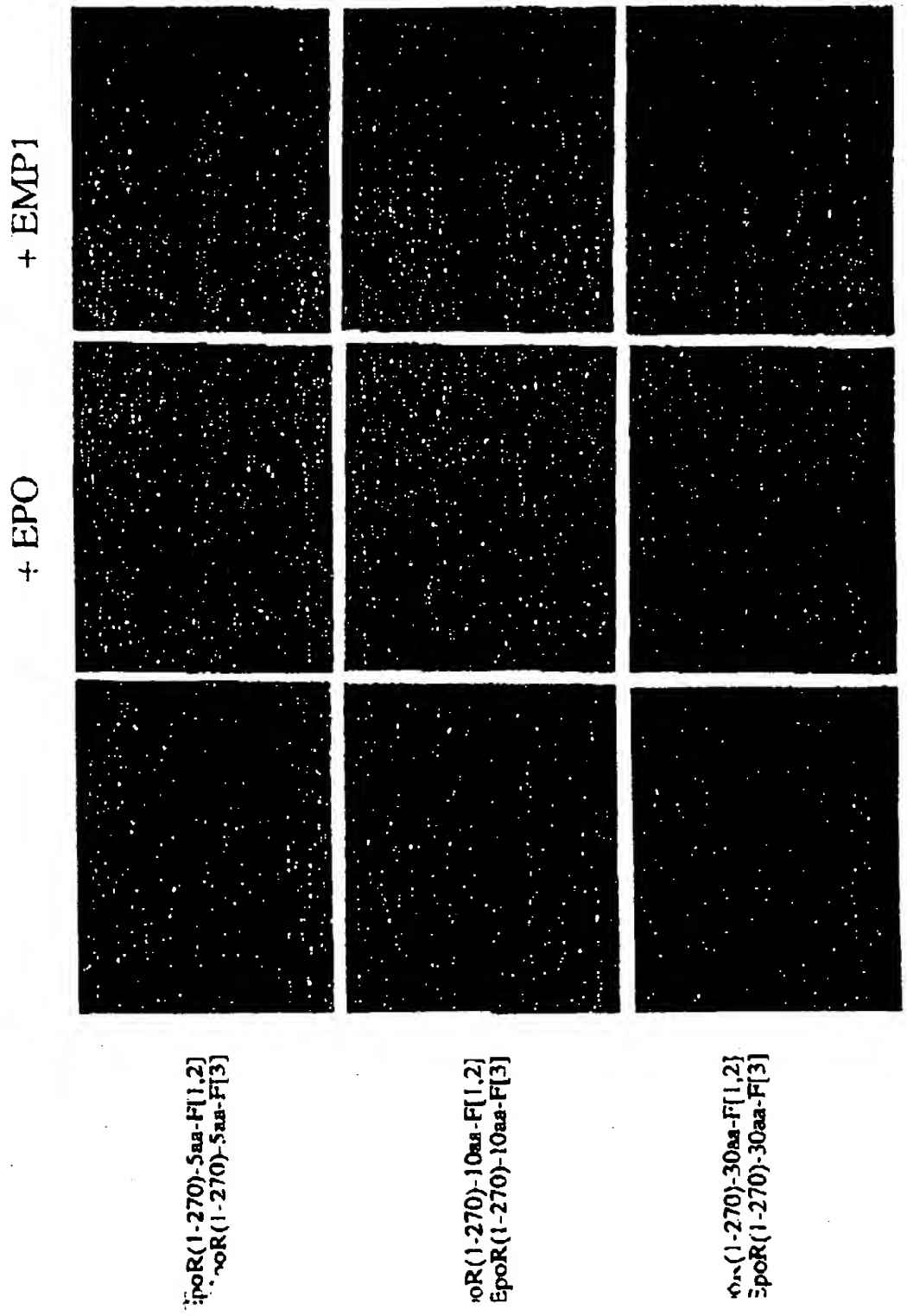

+ EMP1
+ EPO
EpoR(1-270)-5aa-F[1,2]
EpoR(1-270)-5aa-F[3]
EpoR(1-270)-10aa-F[1,2]
EpoR(1-270)-10aa-F[3]
EpoR(1-270)-30aa-F[1,2]
EpoR(1-270)-30aa-F[3]

FIGURE 10A

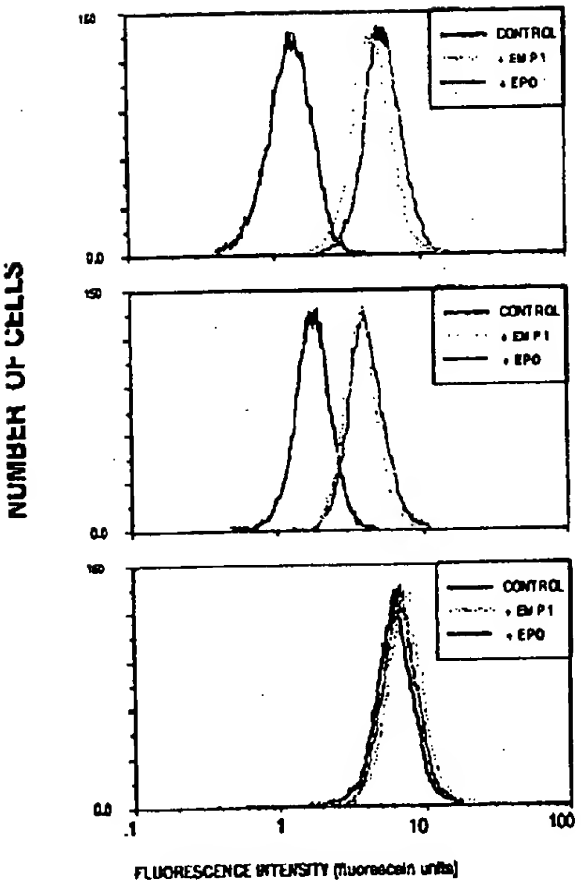

CONTROL
+ EMP1
+ EPO
NUMBER OF CELLS
0.0
.1
1
10
100
FLUORESCENCE INTENSITY (fluorescein units)

FIGURE 10B

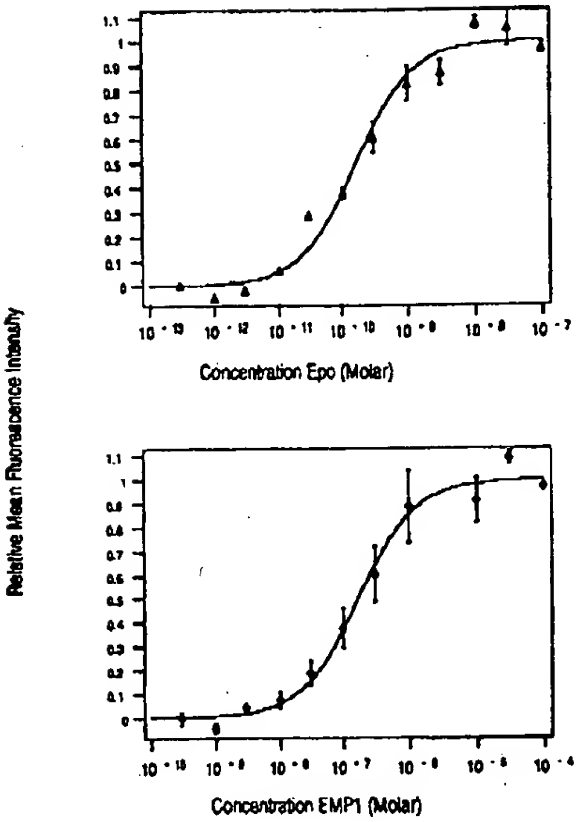
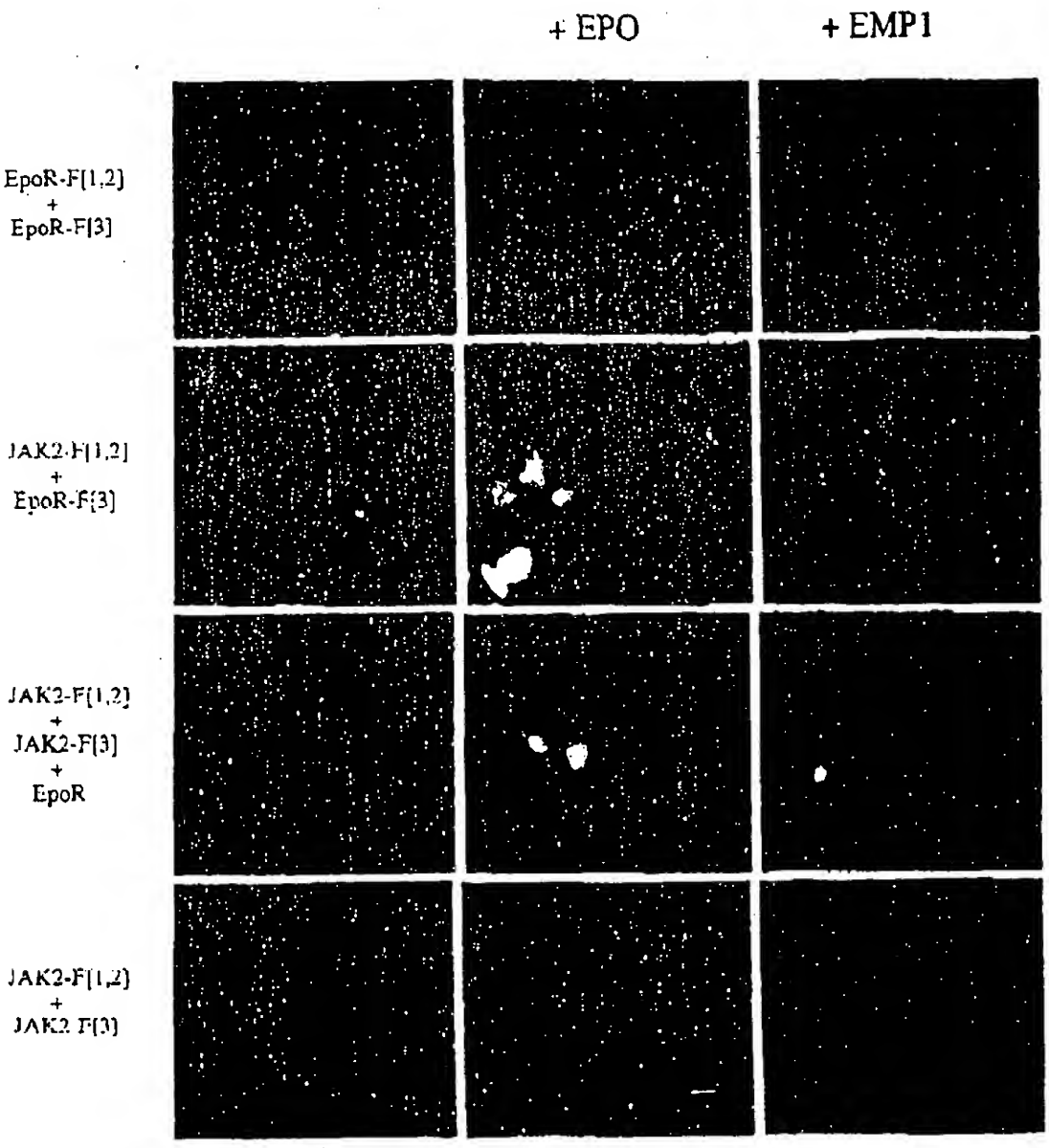

Relative Mean Fluorescence Intensity
Concentration Epo (Molar)
Concentration EMP1 (Molar)

Figure 11


+ EPO
+ EMP1
EpoR-F[1,2]
+
EpoR-F[3]
JAK2-F[1,2]
+
EpoR-F[3]
JAK2-F[1,2]
+
JAK2-F[3]
+
EpoR
JAK2-F[1,2]
+
JAK2-F[3]